Acoustic signal processing with robust machine learning algorithm for improved monitoring of particulate solid materials in a gas flowline

Aminu, Kuda Tijjani; McGlinchey, Don; Cowell, Andrew

Published in:
Flow Measurement and Instrumentation

DOI:
10.1016/j.flowmeasinst.2018.11.015

Publication date:
2019

Document Version
Author accepted manuscript

Acoustic signal processing with robust machine learning algorithm for improved monitoring of particulate solid materials in a gas pipeline

The flow of particulate solid materials in a gas pipeline can significantly erode mechanical equipment, and hence, real-time quantitative monitoring is a timely need for the oil and gas industry. Although a considerable amount of research has been conducted employing acoustic signals for qualitative monitoring, there is still an unmet demand for a simple and robust quantitative monitoring system. Acoustic signal processing with machine learning is a simple and robust method that has the potential to meet this demand, but has not been previously used for particulate solid material quantitative monitoring. Here we report on the development of acoustic signal processing methods strictly on the existence and the significance of the correlation between emitted acoustic signals and the flow conditions and behaviours of particle-laden gas pipeline. The integrated, conventional Artificial Neural Network (ANN) models are used to capture the distribution of the acoustic feature vectors extracted from the signal processing techniques. The backpropagation learning method coupled with Grey wolf optimiser is used to adjust the weights of the network to minimize the regularized cost function for each feature vector. The Grey wolf optimiser is used to provide global adaptation strategy for the network hyper-parameters. The results from the signal processing techniques demonstrate a significant qualitative association between flow conditions and the emitted acoustic signature. Further, conventional ANN has mainly been concerned with capturing systematic patterns in a distribution of measurements fixed in time and the results of the processes are collected in discrete time intervals. Therefore, a modification of the classical ANN, called the Time Delay Neural Networks (TDNN) is used to capture such dynamics. The proposed method compares the performance of the classical ANN models with the TDNN models wherein the feature vectors were used to train the TDNN models. Results show that the TDNN models outperform the classical ANN models which confirm the fact that classical ANN models are insufficient for processing these time sequences. Overall, this study lays the basis for employing signal processing techniques in the development of a real-time quantitative particulate solid monitoring in a gas pipeline.

Keywords:

Acoustic signal processing, Grey Wolf Optimiser (GWO), Neural Networks, Time-Delay Neural Networks (TDNN), condition monitoring

1 Introduction

In general, particulate solid material flow in a pipeline with hydrocarbon fluid remains one of the key technical challenges to the operators in the oil and gas industry. It can cause damage to equipment through erosion, increasing the risk to maintaining mechanical integrity of the equipment and additional costs due to expensive parts replacements and machine downtime[1][2]. In recent years, the increasing trends towards increased oil and gas production particularly in the power industry [3] for continued power generation has increased operating pressures and, therefore, has resulted in more particulate solids in the pipeline. Hence, real-time monitoring of particulate solid material presence in the flow path is a timely need. The aim of implementing a quantitative monitoring system is to achieve uninterrupted production (flow assurance) and curtail its adverse consequential effects. However, most recent techniques for monitoring solid particulate material in hydrocarbon pipeline rely on visualisation of acoustics emitted by solid particles impacting on the walls of the bend component where the sensors were mounted. These methods only provide qualitative information about the flow conditions in the pipeline and cannot provide information on how much solids particles

are in the pipeline. Hence, it is a significant requirement to have a system for real-time quantitative monitoring of solid particles that can provide quantitatively, the flow conditions in the pipeline.

Torbjoern, H. [4] presented a comprehensive review on the lessons learned from sensing and online monitoring of particulate solid materials in a production pipeline using acoustic sensors - from the many years of operation. Other earlier research articles provide evidence of significant progress in this area [5][6][7][8], but still demonstrate the current need for reliable techniques of flow condition monitoring system. However, surface monitoring using 'point' sensors has been the most popular, to date, the use of fibre optic Distributed Acoustics Sensors (DAS) is common [1][9]. Although the traditional monitoring system provide a delayed indication of onset of particulate solid materials events and do not provide sufficient information about the zones that are producing the effect, DAS has been considered as an alternative technology for downhole particulate solid materials detection. Fibre optic DAS are acoustic sensing systems that employ the backscatter component of the light injected into an optical fibre for detecting acoustic perturbations along the length of the fibre. It monitors changes in the length and index of refraction of the fibre caused by impacting acoustic signals. In effect, the fibre works similar to a distributed array of microphones broadcast over the entire length of the borehole sensing downhole acoustics[1]. Despite this research, there is still an unmet demand from the oil and gas industry for a robust quantitative monitoring system which is capable of providing reliable data in the critical area of flow assurance.

Acoustic signal processing with integrated machine learning algorithm has a higher potential to meet this demand. In fact, acoustic signal processing techniques developed for flow conditions recognition is characterised as being knowledge-based approach [10]. On this premise, a sample data is used to train models in advance to identify future data by means of optimisation-oriented approaches. Disregarding the above-mentioned characterisation, it is, however, important to note that relevant information for the recognition may be hidden in the raw signal data. Therefore, an intermediary step that converts the "curse of dimensionality" of the signal of interest to information of reduced and predefined lengths is always required. This is known as the feature extraction step and the resulting operations are called feature vectors. The direct significance is that, the more the features preserve the information required from the acoustic signal, the greater are the results of the recognition [10]. Therefore, signal processing serves the purpose of determining the features needed to construct the recognition models and to confirm the models constructed to represent the phenomena. Nevertheless, the signal data, most likely will have a particular problem that require specific speciality in the analysis.

Real world signals are represented not over an information known a priori and are generally represented based on their probabilistic statistics which can vary over time. As a result, such signals possess nonlinear and nonstationary characteristics. The case is true for acoustic signals where the instantaneous frequency and hence the spectral content changes within the signal duration[11]. Recent trends in research of nonstationary signals reveal that time or frequency domain only analysis are not sufficiently adequate to describe the characteristics of these signals due to the time-varying nature. Therefore a joint time-frequency analysis would be a better approach in an attempt to adequately process the signals. Traditionally, spectral analysis based on Fourier technique has been generally applied to all kinds of signals to provide global information concerning the spectral content of the signals. This is partly due to its simple mathematical prowess. However, despite its general validity, there are some caveats: the signals must be strictly stationary; and the system represented must be linear; otherwise the time occurrences of the spectral information will remain unclear[11][12]. Most methods aimed at processing nonstationary signals are in one way or the other relying on the application of Fourier analysis. Such methods, however, can have some additional shortcomings. For

example, the spectrogram (Short Fourier Transform) is too coarse due conflicting requirements in localizing an event in time or frequency domain; the wavelet analysis is not adaptive being sensitive to gradual frequency changes; the evolutionary spectrum requires a unique technique to define the basis function a posteriori[12]; the Wigner – Ville distribution often introduce negative energy for some frequency ranges[13].

In order to overcome these challenges and effectively process the acoustic signal, feature representations of audio signals previously developed for music and speech processing applications were needed to describe the evolution of the signal in the time and frequency domain. The key to this approach is to presume that on short time scales (usually in the range of milliseconds), the signal is stationary. Although this assumption has provided an important basis for the feature extraction process, its justification in nonstationary signals is not always readily available[14]. In this study, several techniques were developed/employed for signal processing based on discrete fourier transform analysis (DFT), cepstrum and Hilbert Huang Transform (HHT). Additionally, Artificial Neural Networks (ANNs) have been used in the decision making phase. However, there is a lack of techniques in classical ANN which consider the temporal dynamics in the feature vectors. This point is crucial impediment to the performance of the classical ANN. Instead, a time delay neural network (TDNN) is used in the decision making phase.

This paper provides the first step towards the aim to develop a real-time quantitative solid particulate material monitoring system. Glasgow Caledonian University's multiphase flow loop facility was used for the experiments. The concentration and the velocity of the solids were varied during the experimental trials. Conventional contact microphone mounted externally to a pipeline bend was used for recording the emitted acoustic. The output from the sensor processed using different signal processing techniques and the observations are discussed in correlation with the concentration of particulate solid materials that are quantitatively observed in the pipeline via several measurements.

The overall contribution of this study can be highlighted as follows:

i. Development of signal processing methods to study flow behaviours and conditions in particle-laden gas pipeline wherein to the author's best knowledge, no work in the literature has examined the use of such techniques for quantitative monitoring.
ii. Application of classical ANN with swarm intelligence (Grey wolf optimiser) based optimisation method for improved performance of the learning algorithm and then onto TDNN, to capture the array of features extracted from i.
iii. The new metric, normalized root mean square error is introduced to compare the performance of the different flow pattern condition recognition models.

The remainder of this paper is organised as follows. Section 2 explains the experimental setup, materials, and the data acquisition setup used for the experiments respectively. Furthermore, details of the experimental measurement procedure are presented in the section. Section 3 presents the development of the signal processing techniques and the details of the feature extraction methods used as preprocessing step. Section 4 describes the classical ANN in general and the TDNN. The results obtained by the application of the proposed techniques and the machine learning models are presented in section 5. Section 6 provides a discussion of the results and the concluding remarks drawn from the studies and future work are in section 7.

2 Experimental outline

This experiment was designed to determine if acoustic signal processing could be employed to detect the concentration of solid particulate materials in a particle-laden gas pipeline. The experimental objectives of the experimental design were two fold; first closely simulate the real reservoir conditions by injecting concentrations of particulate materials into the pipeline; second generate a series of different solids concentrations in the pipeline.

2.1 Experimental set up

The experiments were conducted using the GCU multiphase flow loop facility, as shown in Figure1. The pipeline layout in the loop was of mild steel internal diameter approximately 50 mm, with total length of around 12 m. The majority of the pipeline layout was in the horizontal plane and included four 90 degree standard radius bends. The acoustic recording sensor was positioned at the bend. The tests were carried out with flow conditions having solid concentrations in the order of 3 to 44 by mass. The gas flow to a sand feeder and transport line was achieved using a nozzle bank and an upstream pressure control valve. The mass flow rate of solids was measured using the change in mass signals from the load cells at the reception tank. The location of the acoustic sensor is illustrated in Figure 2.

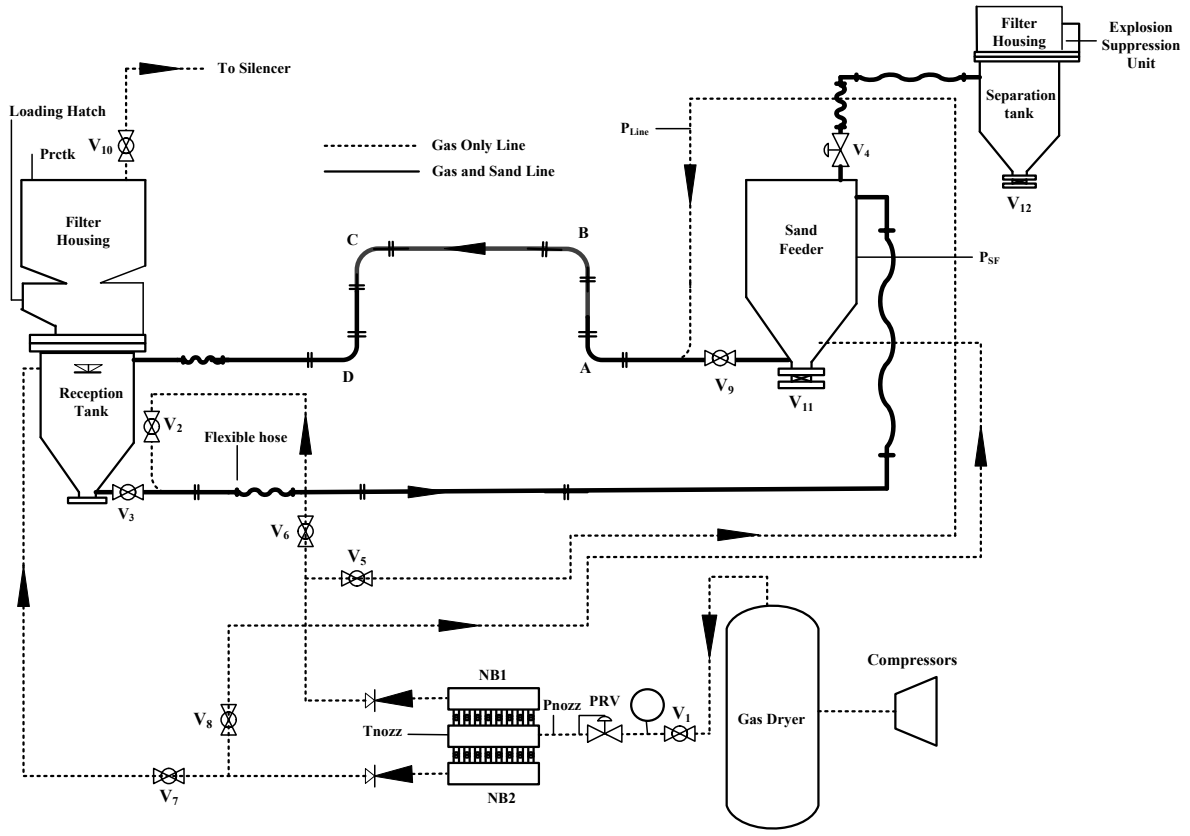


Figure 1 Multiphase flow loop schematic diagram

2.2 Materials

The solid material used for the experiments is the Garside 2EW sand. This sand has an average particle size of 360.18 μm and its size distribution spread between 60 μm and 2000 μm. In this case, the gas used is a compressed air supplied by two screw compressor. The compressors were capable of delivering approximately 0.123 m^3/s of free air with maximum steady pressure of around 7.5 barg.

2.3 Data acquisition setup

The signals emitted by different solid concentrations contribute to the acoustics employed in the preprocessing stage. These signals have been conditioned by a signal conditioning circuit and thus have sufficient signal strength. The complete setup consists of: four contact microphones mounted on the four bends; four single ended pressure transducers to record the pressure at various locations within the layout; a temperature sensor mounted upstream to record the gas temperature; three load cells installed on the reception tank to measure the solid mass flow rate; national instruments USB-6211 and USB-6212 data acquisition systems – UBS-6212 for the microphones and USB-6211 for the other sensors; HP laptop with National Instruments Signal Express software.

The microphone location presented in Figure 2 was chosen based on the assumption that the solid particles provide the highest kinetic energy at the bend location compared to the straight section. Data was collected at a rate of 44100 Hz.

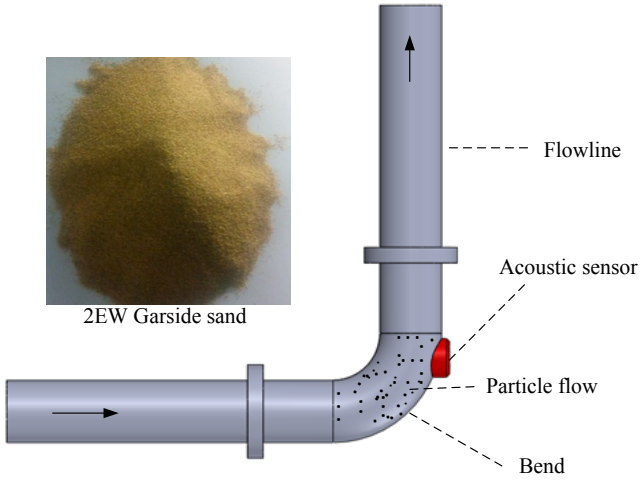


Figure 2 Location of the Acoustic sensor

2.4 Methodology for the experimental measurement

The procedure implemented to measure the solid concentration, solid flow rate, gas velocity and the total line pressure drop is discussed in this section. Sand concentration is determined using equation (1). In equation (1), $\dot{m}_s$ denotes the mass flow rate of sand and $\dot{m}_g$ denotes the total mass flow rate of gas. Figure 3 illustrates different sand concentrations in the pipeline.

$$sand\ Conc = \frac{\dot{m}_s}{\dot{m}_g} \quad \textbf{(1)}$$

However, the sand flow rate is calculated from the test data at steady line pressure using Equation (**2**)

$$\dot{m}_s = \frac{sand\ mass\ Collected\ (kg)}{Time\ (s)} \quad \textbf{(2)}$$

The gas at the transport line can be determined during the test run using the gauge pressure at the supplementary and the temperature measured upstream of the nozzle bank. Since the pressure acting on the sand particles in the transport line is in absolute unit, then the Equation (3) can be used to calculate the gas velocity.

$$V_{gas} = \frac{\dot{m}_g}{\frac{(P_{guage} + 1.01325\ bar)}{RT}} \cdot A \quad \textbf{(3)}$$

The total transport line pressure drop is assumed to be pressure difference between the transport line and the reception tank presented by Equation (4)

$$\Delta P_{Line} = P_{Transport\ Line} - P_{Reception\ tank} \quad \textbf{(4)}$$

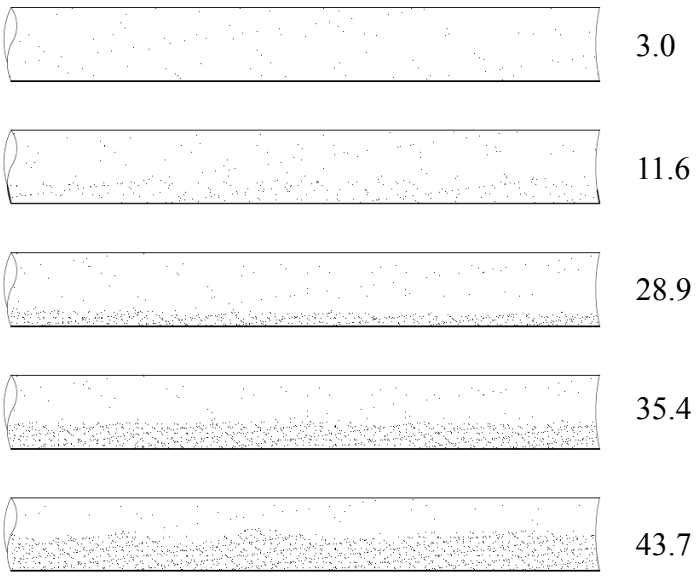


Figure 4 Different sand concentrations in the pipeline

3. Signal Processing

The flow behaviour and flow conditions of particle-laden gas pipeline would, in general, generate acoustics as the particles hit the bend component of the pipeline. Hence, several signal processing techniques were employed to extract relevant features from the acoustic signal in order to study the association between the solid characteristics data in the pipeline and the acoustic emissions. The following section presents the development of feature the extraction techniques.

3.1 Structure for feature extraction techniques

On the basis of nonstationary characteristics of the acoustic signals, it is therefore presumed that on short-time scales the signals are stationary. Based on this assumption, the feature extraction process was decomposed to form a structure for sequential mathematical operations that could lead to a more compact feature computation. The basic outline of the structure can be seen in Figure 4.

Figure 4: Sequence for feature extraction process

The first component of the structure is the block for pre-processing of the acoustic signal. Here the original signal is first decomposed into a small number of frames in sequence. In practice, frame duration of 10 – 30 ms is usually selected to provide a meaningful frequency and time resolutions. Although there is always uncertainty in short -time processing, this interval has been found to be practical[15]. Next, windowing function is applied on each frame to reduce spectral distortions due to discontinuity at the edges of the frame[16]. Many windowing functions have been developed for truncating and shaping signal frames for spectral analysis, but by far, the hamming window is the most common one. It is simple interms of implementation, computationally efficient and guarantees good result[14]. The hamming window $h(n)$, is defined as

$$h(n) = \begin{cases} 0.54 - 0.46cos\left(\frac{2\pi n}{L-1}\right), & 0 \leq n \geq L-1 \\ 0 & otherwise, \end{cases} \quad (5)$$

where L denotes the frame length. This window function simply shrinks down the frame amplitudes to zero at the boundaries. Normally, the windows are time shifted to tackle the problem of significant information losses near the edges of the frame. In particular, if the signal is analysed in frames, information lost at the edge of a frame can be collected between the confines of the following frame. Furthermore, the features extracted following this short-time processing are commonly called short-time features [17][14].

Signal transforms are used to map the representation of the signal from one domain to another. The goal is to change the original interpretation of the signal into a form that could be useful for feature extraction processes. Although different signal transforms can be used to accomplish such mapping, however, the most relevant ones are presented here as follows.

3.1.1 Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) is an increasingly important concept in digital signal processing applications including frequency analysis of signals. This is not surprising, then, given the existence of computationally efficient and powerful algorithms for its evaluation. And it is used to provide frequency domain representation of signals[18][19]. The theme of spectral features is essential and has considerable impact on characterizing and analysing the acoustic contents. For this reason, understanding the mathematical structure of DFT is critical. To obtain the frequency distributions of a given discrete –time signal $\{x(n)\}$, the signal is first sampled in the frequency domain to construct the Fourier transform sequence $X(w)$. However, $X(w)$ is a continuous function defined at infinite frequencies and therefore is not numerically computable. Instead, the Fast Fourier Transform (FFT) algorithm is commonly used to exploit the numerical challenges and efficiently compute the DFT and its inverse.

Consider a discrete-time signal $\{x(n)\}$, of length N defined by $0 \leq n \leq N-1$: $\{x(n)\} = 0$ outside the interval, then the DFT sequence is given as:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{\left(-j\frac{2\pi}{N}kn\right)}, \qquad 0 \leq k \leq N-1, \quad (6)$$

where $X(k)$ is a sequence of frequency samples obtained by evaluating the Fourier transform $X(w)$ at N equally spaced discrete frequencies. Recall the Euler's relationship: $e^{j\omega t} = \cos\omega t + j\sin\omega t$ and

the implication of this is that the outputs of the DFT are complex variables. Thus the magnitude and hence the energy of each of the frequency bin in the spectrum can be obtained by multiplying the respective complex variable by its conjugate. However, in order to recover the original signal sequence $x(n)$ from the frequency samples, the relation

$$x(n) = \frac{1}{N}\sum_{n=0}^{N-1} X(k)e^{\left(j\frac{2\pi}{N}kn\right)}, \qquad 0 \le n \le N-1, \tag{7}$$

is used and is called the inverse DFT (IDFT). Clearly, if the signal sequence $x(n)$ is sampled at a rate f_s, then the bin frequency in hertz corresponds to $f_k = k\frac{f_s}{N}$, $0 \le k \le N-1$ with minimum separation between successive frequency bins of $\frac{f_s}{N}$. The term $\frac{f_s}{N}$ is known as the frequency resolution of the DFT. For a given sampling frequency, if the desire is to have better frequency resolution, then the sequence $x(n)$ can simply be expanded by appending zeros to it, that is, zero padding. It should be pointed out that zero padding does not alter the original form of the sequence but simply refines the sampling of the spectrum already in place [18].

3.1.2 Hilbert-Huang Transform (HHT)

Unlike the Fourier transform, the Hilbert transform provides the instantaneous frequency energy distributions with time localities[12]. It is, therefore, an orthogonal and adaptive time-frequency analysis method for nonlinear and nonstationary signals. The HHT involves two steps. First, the empirical mode decomposition (EMD) technique is applied on the signal from which the local embedded oscillations known as the intrinsic mode functions (IMFs), and a residual understood as the signal trend are extracted. Once the IMFs are extracted, then the next step is to apply Hilbert transform on each of the IMF component. Following this, the Hilbert spectrum is obtained. The empirical decomposition is an iterative procedure known as the sifting process and is based on the simple assumptions: the signal has at least one maximum and one minimum; and that the time scale property is defined by the time reversal between extrema. More importantly, each of the resulting IMFs must satisfy two requirements: on the whole signal, the number of extrema and the number of zero crossings must either equal each other or differ at most by one; and at any point, the average of the envelope defined by the local maxima and that defined by the local minima is zero. The first requirement is analogous to the conventional narrow band conditions for stationary Gaussian process. The second requirement qualifies global condition to a local one, which is needed so that the instantaneous frequency will circumvent false images introduced by riding waves. The first component of the IMF extracted usually represents the highest frequency constituent of the signal, whereas the lower frequency bands are described by the lower order IMFs [12][20][21].

The systematic decomposition of the signal, often designated sifting process, undergoes multiple analytical steps which are successively described as follows[21][12]. For a given signal $x(t)$, first, identify all the local maxima and local minima. Then interpolate the local extrema via cubic spline to obtain the upper envelope $x_{max}(t)$ and the lower envelope $x_{min}(t)$ of the original signal. Compute the mean $m_1(t)$ of the envelopes using the relation:

$$m_1(t) = \frac{x_{max}(t) - x_{min}(t)}{2} \tag{8}$$

Subsequently, subtracting the mean $m_1(t)$ from the original signal $x(t)$ yields:

$$x(t) - m_1(t) = h_1(t) \quad \textbf{(9)}$$

Classically, $h_1(t)$ will be an IMF component which is likely to satisfy all the conditions of IMF. However, in reality unwanted fluctuations are apparent. This results in new extrema and a shift or exaggeration on the existing ones. Repeating the above procedure, the sifting process k times until the average envelope approaches zero, so that $h_1(t)$ qualifies to be an IMF component of the signal designated as:

$$h_{1(k-1)}(t) - m_1(t) = h_1(t), \quad c_1(t) = h_{1k}(t) \quad \textbf{(10)}$$

It is worth noting that this iterative procedure can in fact remove riding waves and smooth uneven amplitudes. In addition, to preserve the physical sense of the meaningful amplitude fluctuations, a stopping criterion for the sifting process is usually defined. Next, the first residue $r_1(t)$ is obtained via $r_1(t) = x(t) - c_1(t)$ and treated as the new signal. Repeat the sifting process on subsequent residual $r_j(t)$ until no further IMF can be extracted and the results is

$$r_1(t) - c_2(t) = r_2(t), \ldots r_{n-1}(t) - c_n(t) = r_n(t) \quad \textbf{(11)}$$

where $r_n(t)$ represents the signal trend which is monotonic and hence no IMF can be extracted, and c_n is the nth IMF component. Finally, the original signal $x(t)$ can be exactly reconstructed in terms of the linear combination of IMFs and residue:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t)\,. \quad \textbf{(12)}$$

Subsequently, applying the Hilbert transform on each of the IMF gives the following relation

$$C_i(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{c_i(\tau)}{t-\tau} d\tau \quad \textbf{(13)}$$

where P denotes the Cauchy principal value of the integral. With this definition, the complexification of $C_i(t)$ and $c_i(t)$ gives the analytic signal, $Z_i(t)$ as

$$Z_i(t) = c_i(t) + jC_i(t) = a_i(t)e^{j\theta_i(t)} \quad \textbf{(14)}$$

The instantaneous amplitude $a_i(t)$ and the instantaneous phase $\theta_i(t)$ are determined from:

$$a_i(t) = \sqrt{c_i^2(t) + C_i^2(t)} \qquad \theta_i(t) = tan^{-1}\left(\frac{C_i(t)}{c_i(t)}\right). \quad \textbf{(15)}$$

Accordingly, the instantaneous frequency of the analytical signal can be obtained from the first derivative of the phase $\theta_i(t)$

$$f_i(t) = \frac{1}{2\pi}\frac{d\theta_i(t)}{dt}. \quad \textbf{(16)}$$

In terms of the instantaneous frequency, the original signal $x(t)$ can be expressed as

$$x(t) = Re\left(\sum_{i=1}^{n} a_i(t)e^{j2\pi \int f_i(t)dt}\right) \quad \textbf{(17)}$$

In addition, the energy of each IMF E_i and the total energy E of the signal $x(t)$ are calculated from the equation

$$E_i = \int_0^{\infty} |c_i(t)|^2\, dt \qquad E = \int_0^{\infty} |x(t)|^2\, dt \quad \textbf{(18)}$$

The energy contribution of each of the IMF, EC_i, is therefore obtained from

$$EC_i = \frac{E_i}{E} \tag{19}$$

3.1.3 Discrete Cosine Transform (DCT)

The discrete cosine transform is yet another promising technique widely and extensively used in signal processing applications due to its non-complex nature. This implies a useful property that the DCT coefficients are purely real. Unlike DFT, the DCT expresses a signal as an aggregate of cosine functions of different amplitudes and frequencies. However, despite this distinction, both operate on a signal sequence of finite number discrete data points [18][19].

Given a sequence of discrete-time signal $x(n)$ defined at discrete points N such that $0 \leq n \leq N-1$then its corresponding DCT is given as

$$X_{DCT}(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi k(2n+1)}{2N}\right), \qquad 0 \leq k \leq N-1 \tag{20}$$

However, to reconstruct the sequence $x(n)$ from the generated N-point DCT coefficients, $X_{DCT}(k)$, $0 \leq k \leq N-1$, the inverse discrete cosine transform (IDCT) is used and is given by the following equation

$$x(n) = \sum_{k=0}^{N-1} \alpha(k) X_{DCT}(k) \cos\left(\frac{\pi k(2n+1)}{2N}\right), \qquad 0 \leq k \leq N-1 \tag{21}$$

where

$$\alpha(k) = \begin{cases} \sqrt{\frac{1}{N}} & k = 0, \\ \sqrt{\frac{2}{N}} & 1 \leq k \leq N-1. \end{cases} \tag{22}$$

The DCT has also been popular in other applications including image compression due to its computational efficiency [18]. In this work, DCT is employed to decorrelate sand concentration signal log spectra in mel-frequency analysis so that useful features can be extracted from the signal.

3.1.4 Aggregations

For better a description of feature dynamics, the notion of individual short-term features becomes meaningless, and hence a sequence of short-term features is needed. According to this type of processing, the sequence of short-term features is called the mid-term segment and a signal could have several segments. In practice, however, the duration of mid-term segment typically falls in the range 1 – 10 s and is based on the assumption that within the interval, the segments exhibit like behaviour[15]. Therefore aggregation of features over each of the segment using simple summary statistics (such as mean, median, maximum and minimum) can capture the dynamism of features that best describe the characteristics of the signal. Moreover, it can enhance classifier performance over directly working on the short-term features. The reason is that the series of feature values in the sequence is being mapped onto a scalar value. Hence aggregation on its own can be seen as an important element in the data reduction procedure. In general, the successful execution of the just described structure for feature extraction could result in a feature vector that is compact and complete with high discriminant power[17][22].

Often, in the process of extracting features, it is the domain of the feature that determines the interpretation of the feature data and provides a useful indication about computational requirements. The domain of the feature is the depiction that a feature exists in after extraction. Therefore, the r the time and frequency domains are considered in the acoustic feature extraction process.

3.2 Time-domain features extraction

In the application of acoustic signal prediction, the time-domain is important; it describes the signal waveform directly as well as the features extracted. Typical examples of short-term features in this domain include: energy and zero-crossing rate. Although time-domain features provided meaningful information about the signal, it is important to note that they are not sufficient for characterization of the signal. Instead a combination of them with the state-of-the- art frequency domain features is used to provide the needed characterization for ensured predictive accuracy. Additionally, the time-domain is the basis from which features from other domains are derived from the signal. The sections that follow attempt to briefly describe some of the most commonly used time domain features.

3.2.1 Short-Term Energy

According to the classical definition, the energy of a signal describes its capacity to perform work [10]. In order to apply the concept of energy in the analysis of sand concentration acoustic signal, it is necessary to bring the notion of 'impingement' into considerations so that the concept of energy is applicable. Thus the energy of sand acoustic signal relates to the impingement of the sand particles with the bend, which could be due to particle concentration, flow velocity and gas pressure [23]. All these factors tend to significantly affect the signal amplitude which in turn dictates the energy of the signal. The short- term energy of the signal is given by

$$E(i) = \sum_{n=0}^{W_L-1} x_i(n)^2 \,, \tag{23}$$

being $E(i)$ and x_i, respectively, the energy and the sequence of amplitudes of the i^{th} frame, with W_L representing the width of the window. For convenience, the energy is normalized to eliminate dependency on the frame width by simply dividing it with W_L , and the relation becomes

$$E(i) = \frac{1}{W_L} \sum_{n=0}^{W_L-1} x_i(n)^2 \,. \tag{24}$$

Based on equation (23), the short-term energy can be defined as the mean energy per frame which is actually considered as a measure of frame signal power with respect to sand particles impact[24]. Similarly, using equation (21), the energy contribution of the first - two higher frequency constituent of the signal is calculated and included in the response feature vector.

3.2.2 Zero-crossing rate (ZCR)

For a given signal frame, zero crossing rate is a weighted time – domain crossings within the frame i.e the number of times the signal changes sign within the frame [25] and its mathematical definition is

$$ZCR(i) = \frac{1}{2W_L} \sum_{i=0}^{W_L-1} |sign[x_i(n)] - sign[x_i(n)]| \tag{25}$$

and it follows that

$$sign[x_i(n)] = \begin{cases} 1, & x_i(n) \geq 0, \\ -1, & x_i(n) < 0. \end{cases} \quad \textbf{(26)}$$

The ZCR is simple and cheap in terms of computational complexity. Conventionally, the ZCR correlate with the spectral behaviour of the signal under analysis [10]. This is true for narrowband signals, but for broadband signals like the sand concentration signals the interpretation of zero crossing rate is much less accurate and therefore approximate spectral characteristics can be obtained using representation based on the short-time zero crossing rate. In addition, the ZCR can also be interpreted as a metric for randomness [15].

3.2.3 Energy Entropy

The energy entropy is a feature used to detect sudden changes in the energy level of a signal due to sand particles/inner wall/gas interactions at the bend. For the computation of the energy entropy[15], the signal frame is distributed among *K* sub-frames of equal duration. Next, the energy contribution of each sub-frame, e_j , $j = 1,2, \dots, K$, is computed using the relation given in equation (29). The idea is to create a sequence of *K* independent variables within the signal frame defined by a certain probability distribution.

$$e_j = \frac{E_{sub-frame\ j}}{E_{frame}}, \quad \textbf{(27)}$$

with $E_{sub-frame\ j}$, denoting the energy of j^{th} sub-frame and E_{frame} , representing the total energy of the frame. Finally, the Shannon entropy[26] is applied on the sequence in an attempt to characterize the unpredictability of the frame signal energy. For the sub-frame energy sequence, e_j, the energy entropy is obtained by:

$$EE(i) = -\sum_{j=1}^{K} P_j(e_j)\ \log_2\left(P_j(e_j)\right) . \quad \textbf{(28)}$$

By equation (30), P_j is the probability of the j^{th} component in the energy sequence and has been estimated by dividing j^{th} component in the sequence with the total frame energy. Conceptually it is interpreted as a measure of uncertainty/disorder in a system for larger values[27]. Physically, it has meaning on the detection of sudden changes in the energy magnitude within a signal frame for lower values. The intuition is that, if one of the sub-frames gives a high energy value, then it implies that the probabilities will not be equal and the entropy is decreased resulting into a lower value[15].

3.3 Frequency-domain features extraction

Understanding the spectral structure of a signal will provide a strong basis for characterization among the different responses. However, to uncover the spectral distribution of a discrete-time (sampled) signal, the DFT has become the most commonly used algorithm due to its representational and computational advantages. The output of the DFT simply provides the frequency bins present in the signal and associated energies following which the spectrum is obtained. Sequel to that, features that give important information about the spectral characteristics of the signal are computed from the spectrum. Therefore, we give, as follows, a brief description of the so-called spectral features that have proved meaningful in practice when analysing the spectrum of signals purely for discriminatory purposes. Note: for simplicity and notational convenience, let W_{fL} be the number of frequency bins (similar to the numeric samples of W_L in the time domain) in the mathematical representation of the features.

3.3.1 Spectral Centroid (SC)

The spectral centroid of an acoustic signal frequency spectrum reflects the center where most of the energy in the frequency distribution is concentrated. It indicates whether the spectral structure of the signal contains a generality of low or high frequencies, respectively [28][29]. The spectral centroid of the i^{th} signal frame is computed as seen in equation (31)

$$SC(i) = \frac{\sum_{k=0}^{W_{fL}-1} k.|X_i(k)|}{\sum_{k=0}^{W_{fL}-1}|X_i(k)|} \quad \textbf{(29)}$$

where k is the frequency bin for the corresponding frame. Obviously, a large SC values signifies that the spectrum is dominated by high frequencies whereas a small SC values shows that the spectrum is dominated by low frequencies.

3.3.2 Spectral Spread (SS)

Unlike the spectral centroid, the spectral spread provides a metric about the shape of the spectrum. In other words, is a measure that is used to describe whether the spectrum is concentrated in the proximity of its centroid or it is simply disperse over the spectrum[15][17]. Its mathematical definition is given in equation (32)

$$SS(i) = \sqrt{\frac{\sum_{k=0}^{W_{fL}-1}\big(k - SC(i)\big)^2 X_i(k)}{\sum_{k=0}^{W_{fL}-1}|X_i(k)|}}\,. \quad \textbf{(30)}$$

Though the signal has broadband instinct, still the spectral spread can provide meaningful correlation between different responses even when they have different power spectrum. In practice, the spectral centroid and the spectral spread are usually normalized in the range [0, 1], simply by dividing their resulting values with $\frac{f_s}{2}$.

3.3.3 Spectral Entropy (SE)

The spectral Entropy (SE) is a feature used to capture the degree of irregularity in the signal spectrum. It is also a measure that relates with the peakiness/flatness of the spectrum[30][31]. From the computational perspective, the spectrum of the short-term frame is divided into K non-overlapping sub-bands. The spectrum in each of the j^{th} sub-band, $j = 0, \dots, K-1$ is then normalized by the full-band spectrum to obtain the probability function. For the sub-band normalization, the relation in equation (33) is utilized [32].

$$s_j = \frac{S_j}{\sum_{j=0}^{K-1} S_j}, \quad \textbf{(31)}$$

being S_j and s_j the energy of the j^{th} sub-band and the probability function (i.e the normalized sub-band spectral energy) respectively. The entropy is then computed using the normalized spectral energy s_j according to the equation:

$$SE(i) = -\sum_{j=0}^{K-1} s_j \,.\log_2\big(s_j\big). \quad \textbf{(32)}$$

Nevertheless this sub-division, equation (34) gives the full-band entropy for the signal frame [32]. Its value is low for a flat frequency distribution and high for a spectrum with sharp peaks.

3.3.4 Spectral Flux (SF)

The spectral flux is a basic measure for sudden detection of changes in the signal spectral structure overtime. That is, the spectral flux captures the variation of the spectrum between two adjacent frames in the signal. Mathematically, it is the normalized difference vector of the frame-to-frame spectral magnitude [33][34] given by

$$SF(i) = \sum_{k=0}^{W_{fL}-1} \left\{H\left(X_i(k) - X_{i-1}(k)\right)\right\}^2, \quad \textbf{(33)}$$

where $H\left(X_i(k)\right) = \frac{X_i(k)}{\sum_{l=0}^{W_{fL}-1} X_i(l)}$, is the *k*th normalized DFT coefficient at the *i*th frame. In general sand concentration signals tend to exhibit slowly varying spectral characteristics.

3.3.5 Spectral Rolloff (SR)

Spectral Rolloff (SR) is also a simple feature that is extensively used in a variety of machine learning applications. For instance Marko [35] use it to classify speech acoustic signal. It is defined to be C% percentile of the power – frequency distribution. In other words, it is the percentage below which the frequency bins of the magnitude distribution of the spectrum is accumulated [17][15]. That is, how high in the signal's spectral structure a certain portion of the energy lies. Its mathematical definition can be seen in equation (36)

$$SR(i) = C \cdot \sum_{k=0}^{W_{fL}-1} |X_i(k)|^2 , \quad \textbf{(34)}$$

where C is the percentile value typically in the range 85% - 95%, $X_i(k)$ is the frequency magnitude of the signal at the *i*th frame and W_{fL} is the number of frequency bins. The characteristics of low concentration signals tend to have a lower SR than high concentration signals. High concentration signals contain relatively higher broadband frequencies from sand impacts in the pipeline. The consequence is that high concentration signals tend to have high values of SR. It is worth pointing out that in the actual implementation, the SR value is normalized by the size of the frequency bin.

3.3.6 Mel-Frequency Cepstral Coefficients (MFCCs)

MFCC algorithm is one of the standard techniques for acoustic characterization and is widely used in various domains including bioacoustics identification[36] and ASR [37]. Indeed, the coherency of this standardized algorithm allows the efficient calculation of several cepstral coefficients which are usually taken as features. Additionally, it is perceptually inspired algorithm that mimics the hearing characteristics of the human ear [38]. The resulting coefficients are used to represent the spectral energy distribution of the signal. A rather more complete computational sequence for extracting MFCC features is depicted in Figure 5.

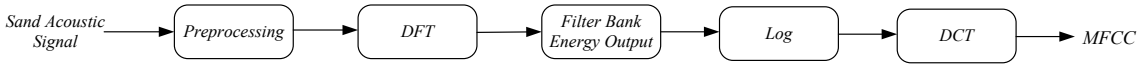


Figure 5: MFCC computation

As shown in figure 4, the first step is to pre-process the signal as described in the previous section. After pre-processing the signal, the DFT is applied to compute the spectrum magnitude which is then warped according to the melodic (Mel) scale using equation (35) in order to accommodate the frequency resolution of the human auditory system. This is achieved using a set of overlapping triangular bandpass filters (known as the Mel filter bank) that lead to an even distribution of the spectrum into a number of critical-band frequencies within the auditory range [35]. More importantly these filters are spaced more densely in low frequency region than high frequency region as indicated in Figure 6.

(a)

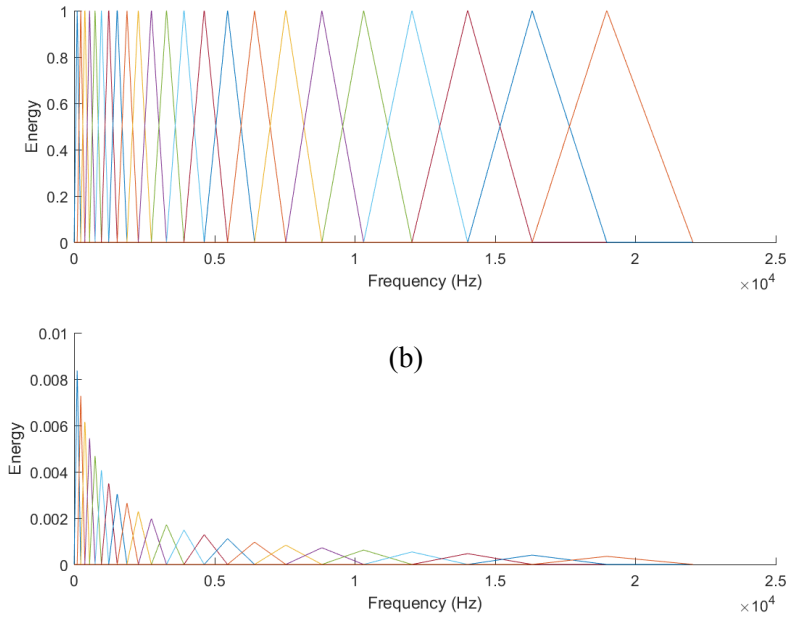


Figure 6 illustration of filter bank frequency responses for the MFCC a) Normalized to unit height b) Normalized to unit area

This Mel filter bank as defined by the critical-band frequencies $f_{cb}(m)$ is given [39] according to the relation in equation (36).

$$f_{mel}(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right), \tag{35}$$

$$W(m,k) = \begin{cases} 0 & \text{for} \quad f(k) < f_{cb}(m-1) \\ \dfrac{f(k) - f_{cb}(m-1)}{f_{cb}(m) - f_{cb}(m-1)} & \text{for} \quad f_{cb}(m-1) \le f(k) < f_{cb}(m) \\ \dfrac{f(k) - f_{cb}(m+1)}{f_{cb}(m) - f_{cb}(m+1)} & \text{for} \quad f_{cb}(m) \le f(k) < f_{cb}(m+1) \\ 0 & \text{for} \quad f(k) \ge f_{cb}(m+1) \end{cases}, \tag{36}$$

where f_{mel} is the logarithmic scale of f linear frequencies. Then the sum of energies of each filter weighted with the corresponding band filter shape is obtained. Subsequently the logarithm function is applied to compute the logarithmic energies of each of the filter and the entire operation is represented by equation (37)

$$E_{m,i} = \log \sum_{k=0}^{W_{fl}-1} W(m,k) \cdot |X_i(k)|^2 \tag{37}$$

where $m = 1, \dots, K$ denotes the filter index, k is the DFT bin index, i is the frame index, $E_{m,i}$ is the logarithmic energy coefficient of the corresponding filter, $W(m,k)$ is the filter shape, and $|X_i(k)|$ is the mel- spectral magnitude vectors whose components are highly correlated. Finally, the MFCCs are obtained by calculating the DCT of the log energies $E_{m,i}$ using

$$c_{n,i} = \sqrt{\frac{2}{K}} \sum_{m=1}^{K} E_{m,i} \cos\left[n\frac{\pi}{K}\left(m - \frac{1}{2}\right)\right], \quad \textbf{(38)}$$

where $n = 1,2, \dots, K$ and $c_{n,i}$ represents the n^{th} MFCC of the i^{th} of the signal. This DCT is usually employed to decorrelate these components thereby improving their statistical properties[37]. In most practical application, the first 8-13 coefficients of the MFCC are suffice to represent the spectral shape of the signal. The implication is that the MFCC is actually evaluated for a number of coefficients that is less than the number of the filters. However, the interpretation of the MFCC coefficients is often a subject of controversy. This is because only the first two coefficients have meaningful interpretation. The first coefficient represents the average power over all frequency bands and the second coefficient provides tradeoff between high and low frequencies in the signal spectrum and often is related to the centre of gravity of the spectrum. The other high-order coefficients contain finer spectral details useful to distinguish the signals of interest with no definite interpretation. Notwithstanding the limitation, they tend to provide compact spectral representation which largely attributed to their success in acoustic identification [17][40].

Other than the above spectral features, there are also some additional spectral features such as the spectral power, the dominant frequency and the dominant frequency magnitude. Features like these, though useful in discriminating between different solid signal characteristics, are not readily available in the literature. Moreover they do not have a definite mathematical definition. As such nothing further can be discussed in relation to these features here. Having described the feature extraction process, the following section will discuss the modelling for machine learning algorithm and the optimisation algorithm.

4 Modelling for machine learning and optimisation algorithms

4.1 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) are computational model inspired in the biological neurons and have been evolved since the seminal neural model of McCulloch and Pitts[41]. More generally, they are used in applications ranging from concept learning to function approximation. Architecturally, they are made up of highly interconnected units called neurons usually organized in layers and aimed to solve a variety of really sophisticated computational problems. This basically consists of an input layer, an output layer and one or more layers between the input and the output often called the hidden layer(s)[42].The connections between the neurons have weights associated them and these weights are random at the initial instant. Moreover, ANNs are categorized into groups according to the connection pattern. Feed-forward networks, in which the connections between layers are unidirectional and recurrent networks which contain feedback connections. However, for simplicity we focus on feed forward ANNs. Although different models have been developed based on the feed forward architecture, the multi-layered feed forward networks are the commonest[43]. Each neuron in the network performs simple computation and application of activation function. This activity serves to generate an input to any of the participant neuron in the succeeding layer[42]. On one hand, the

application of the activation function is especially important in ANNs, for model of complex non-linear classification patterns, the activation functions are crucial. Of the available activation functions, the sigmoid function is far the most common that prevails in ANNs and particularly the logistic function. The sigmoid function exhibits asymptotic behaviour, smoothness, continuous and monotonically increasing characteristics. Others are hyperbolic tangent and arctangent functions[44].

Interestingly, since these weights are random initially, there is the critical need for an algorithm than can modify the weights of the ANN according to a certain well defined rules so that the desired output can be obtained from the network. This process of weight modification is called learning and the method of learning is known as supervised learning. It is supervised because the network is presented with samples that have inputs and outputs. Additionally, the application of supervised learning can be divided into two groups: batch (also known as off-line) and stochastic (also known as on-line). Off-line learning is an approach that utilizes a set of samples in order to learn and approximate a certain function. Conversely, on-line learning is related to adjusting the network weights each time a simple is presented and the error is accumulated for the entire sample size[45][46]. Although there are many learning algorithms for ANNs, the most popular, is adopted here and will be introduced in the following section. This is typically the error correction rule that tends to minimise the learning error through the adjustment of the network weights. More importantly the characteristic of ANNs in the ability to learn exhibit the essential feature of intelligence. Besides the learning capability, the generalization ability of ANN is another important characteristic that strongly influences the performance of the network. Together these factors provide important basis that suggest ANNs are very powerful tools[47].

4.2 Back-Propagation Algorithm (BPA)

Back-Propagation Algorithm (BPA) is still the most widely used learning algorithm for multi-layered feed forward networks. Conceptually, it is simple and computationally efficient[48]. Moreover, the BPA is a gradient-descent based approach commonly employed to minimise the cost function which the network implement in the learning procedure. The sum of squared difference error cost function is the most popular in ANN and is defined in Equation (41) [49]. The simple reason for this is in order to exploit the convex instinct of quadratic equations.

$$J(W) = \frac{1}{2}\sum_{i=1}^{N}\left(d^{(i)} - \hat{y}^{(i)}\right)^2 \quad \textbf{(39)}$$

In equation (39), W denotes the network trainable parameters, d is the desired output and $\hat{y}$ is the network output. In general, the basic principle of BPA is to propagate the error signal of the output layer backward into the network through gradient computation of the arbitrary cost function with respect to each of the connection weights. As a consequence, these weights are iteratively adjusted thereby reducing the error, which serve as a measure for the network performance. Difficulties arise, however, when an attempt is made to design the algorithm for implementation on any real platform. The evidence of these difficulties becomes very clear in making arbitrary choices such as the number of hidden layers, the number of neurons in a hidden layer, the learning rate, the momentum term, the error stopping criterion and the like. The first two reflect on the structure of the network whereas the last three relate to the learning parameters. Although these choices are critical, there are no straightforward approaches for deciding them as they are primarily problem and data dependent. In this respect, the practitioner can make better choices heavily relying on heuristics and some underlying theoretical results [44].

For number of hidden layers, there used to be arguments among researchers about networks with single or more hidden layers – which one provide better mapping characteristics? Theoretically, both architectures can approximate non-linear function to the desired level of accuracy, but the intuition is to first try solving the problem at hand using the conventional ANN (network with single hidden layer) seems reasonable and then evaluate the network's performance [44]. However, the recent trend in research has seen a dramatic shift from the conventional networks to networks with many hidden layers that have led to the introduction of the term deep learning methods. In particular top level of performance had been reported at solving really complex problems involving deep neural networks[50]. The notion of the number of neuron in a hidden layer, on the other hand, is critical in relation to the generalization ability of the network. There are two basic difficulties with the idea of number of neurons in a hidden layer as follows. The first one arises from insufficient number of neurons. Consequently, the network simply neglects the underlying function which results in poor network performance[51][52].The second difficulty arises from too many hidden layer neurons. It is well known that real world observations are composed of data and noise. Therefore instead of modelling just the data, the network also fits the noise. Although the network will perform incredibly well on the training samples, it will not provide reasonable outputs on unknown samples. This phenomenon is known as overfitting[44]. Notwithstanding, this difficulty is no longer a serious issue since the introduction of validation technique in quantitative models. The details will be discussed in later section.

However, since the BPA implements the steepest descent (SD) technique to update the network weights, the appropriate choice of the learning rate, which "scales" the gradient, is crucial and has significant effect on the convergence time. Depending on the curvature of the error cost function, a small learning rate will result in too many steps in order to reach a reasonably good solution. On the other hand, a large learning rate will perhaps lead to oscillation and thus preventing the error to fall below a certain value. Additionally, for stabilization particularly in error functions that have steep ravines landscape, the momentum term is introduced in the weights update rule. The momentum term scales the influence of the previous weight on the current weight thereby reducing rapid oscillations in shallow regions of the error curvature. As a result, working with the momentum term accelerates the speed of optimal convergence[42][53]. Overall, the success of the learning process is dependent on the proper choice of both the learning rate and the momentum term. Their value typically lies in the interval [0,1].

One of the major drawbacks of BPA is convergence to local minima [54]. However, the reality of local minima is a consequence of the fact that the error curvature is simply the superposition of nonlinear activation functions that may exhibit local minima at different locations, which occasionally results in a nonconvex curvature of the error cost function[51][52]. So far, a great variety of techniques aimed to accelerate the learning process have been proposed. These techniques include the Conjugate Gradient BPA, the Broyden-Fletcher-Goldferb-Shanno Quasi-Newton (BFGS) method, Levenberg-Marquardt Algorithm and the like [48]. However, as the basis for these methods is the second order derivative, an obvious difficulty is the occasional convergence to local minima. One way to overcome this challenge is through the use of improved gradient-based algorithms that employ parameter adaptation strategies[43]. This can be achieved by the use of global optimization techniques which can lead to optimal weight adjustments thus allowing the network to eschew local minima during the learning process. Many such algorithms implementation in ANN have been presented in the literature like the Simulated Annealing (SA) [55], Genetic Algorithms (GA) [51], Evolutionary algorithms (EA) [56]. This paper proposes a new methodology for parameter adaptation; "Grey wolf"

optimiser. Therefore, before going any further, it is important to derive the BPA-based learning procedure.

4.3 Derivation of BPA learning procedure

To illustrate the back propagation algorithm, consider the conventional network which has one hidden layer and one neuron in the output layer. The cost function is a function of the training samples and the weights. Therefore the learning proceeds by changing only the weights since the training samples are fixed. Applying the chain rule repeatedly, the gradient of the cost function with respect to each of the weight in the network can be computed and then propagated back through the network, and all the weights are adjusted so as to decrease the cost. The gradient of the output weight vector is

$$\frac{\partial J(W)}{\partial W^{ol}} = -\frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial u_{ol}}\frac{\partial u_{ol}}{\partial W^{ol}}, \tag{40}$$

where W^{ol} is the weight vector for the output layer, $\hat{y}$ is the output and u_{OL} is the weighted sum of the inputs of the output neuron. The negative sign is an indication that weight update is in the direction of negative gradient. From equation (40) the error signal $\delta_{ol}$is

$$\delta_{ol} = -\frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial u_{ol}}. \tag{41}$$

Similarly, the gradient of the hidden layer weight vector is

$$\frac{\partial J(W)}{\partial W^{hl}} = -\frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial u_{ol}}\frac{\partial u_{ol}}{\partial o_{hl}}\frac{\partial o_{hl}}{\partial u_{hl}}\frac{\partial u_{hl}}{\partial w^{hl}}, \tag{42}$$

where W^{hl} is the weight vector for the hidden layer, o_{hl} is the output of the hidden layer neuron(s) and u_{hl} is the weighted sum of inputs of the hidden layer neuron(s). Thus from equation (42) the hidden layer error signal δ_{hl} is written as

$$\delta_{hl} = -\frac{\partial J}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial u_{ol}}\frac{\partial u_{ol}}{\partial o_{hl}}\frac{\partial o_{hl}}{\partial u_{hl}}. \tag{43}$$

Finally, the adjusted weight vector is computed as follows:

$$W^{t+1} = W^{t} + \eta_r \Delta w + \alpha_m (W^{t} - W^{t-1}), \tag{44}$$

where η_r is the learning rate, α_m is the momentum term and Δw is computed using Equation (40) and Equation (42) for the respective layers. However, the weight updation strategy is repeated many times with the training samples until the network converges to producing the desired minimum cost

4.3.1 Regularization

In order to combat the complexity of ANN and avoid overfitting, the regularization technique is employed. This technique constrains the Back propagation algorithm to fit the data well thus improving model performance, particularly when noise is present [57]. The way it works is by shrinking complex model coefficients towards zero values and hence the name weight decay [58]. One way to implement regularization is to introduce the regularization term λ on the model complexity in the cost function. Therefore instead of minimizing the cost function alone, now a combination of the cost function and the model complexity is minimized. In this sense, the regularization hyper-parameter λ is used to trade-off between the cost function and the complexity of the model.

$$J(W) = \frac{1}{2}\sum_{i=1}^{N}\left(d^{(i)} - \hat{y}^{(i)}\right)^2 + \frac{\lambda}{2}\sum (W)^2 \tag{45}$$

Indeed, a proper choice of the regularization term in equation (45) would guarantee a model with optimal complexity will be produced. Together the learning rate, the momentum term, the regularization coefficient are called the network hyper-parameters whose optimal values have to be tuned using the global adaptive technique. Before introducing this technique, we will first outline the structure of the ANN model.

4.4 Modelling using classical ANN

A six layer artificial neural network has been modelled. The six layer ANN shown in Figure (7) has five hidden layers. In this network, the neurons in the layers of the hidden layer are arranged such that the preceding layer has more neurons than the succeeding layer. That way the complexity of the model increases as the data progresses through the network. The processing neurons in the hidden layer and the output layer are nonlinear using standard logistic function. The architecture of the model is thus 2*L* 25*N* 15*N* 10*N 5N 3N 1L*, where L represents linear processing neuron and *N* indicates nonlinear processing neurons. This model is used to capture the propagation of the acoustic feature vector for each class in the training sample.

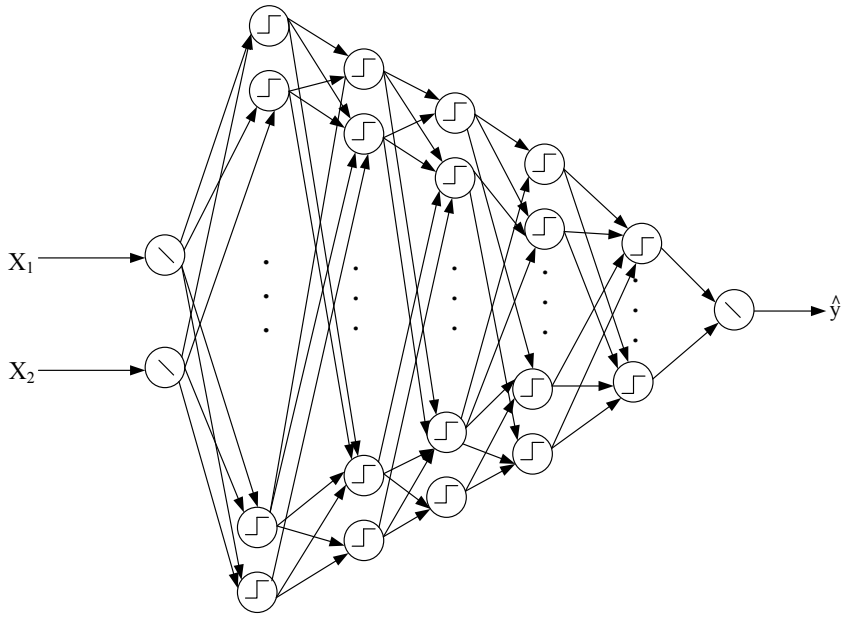


Figure 7: A 6 layer ANN model

The back propagation algorithm is used to update the weights of the network and the Grey wolf optimiser (GWO) adapts the network hyper-parameters while minimizing the regularized cost function for each feature vector. Therefore the inclusion of GWO in BPA is to provide adaptive tuning of the parameters and thus improving the convergence speed of the BPA. We will introduce the GWO needed to adapt the parameters for fast convergence of BPA.

4.5 Grey Wolf Optimiser (WO)

Grey Wolf Optimizer (GWO) is a nature-inspired metaheuristic optimization algorithm firstly developed by Mirjalili *et al* [59] and mimics the social behaviour of grey wolves in a pack. The algorithm is inspired by the leadership hierarchy and hunting strategy of grey wolves. In order to simulate the leadership hierarchy of wolves for algorithmic development, four types

of wolves are considered – the alpha (α), beta (β), delta (δ) and omega (ω). Moreover, the three hunting strategies of wolves: searching for prey, encircling prey and attacking prey are employed. The next subsections present Algorithm development.

4.4.1 Leadership hierarchy

Mathematically, GWO is modelled by assuming the best solution to be the alpha. Accordingly, the second and third optimal solutions are regarded beta and delta respectively. The remainder of the prospective solutions are considered to be omega. The optimization strategy (hunting) in the GWO algorithm is dictated by alpha (α), beta (β) and delta (δ). The omega wolves always submit to the commands of these three wolves [59]–[61]

4.4.2 Encircling prey

As the grey wolves are close to the prey during the hunt, then all the wolves will take position and encircle the prey. The encircling behaviour can be mathematically represented by the equations:

$$D = \left|CX_p - AX(t)\right| \tag{46}$$

$$X(t+1) = X_p(t) - AD \tag{47}$$

where t represents the current iteration, X is the position vector of a wolf whereas X_p denotes the position vector of the prey , A and C represent the coefficient vectors and D is the encircling behaviour. The vectors A and C are random and adaptive vectors that provide exploration and exploitation for the GWO algorithm and are computed as follows:

$$A = 2\,\vec{a} \times r_1 - \vec{a} \tag{48}$$

$$C = 2 * r_2 \tag{49}$$

where r_1 and r_2 are random vectors in [0,1] and the components of $\vec{a}$ varies linearly from 2 to 0 over the iteration circle [59]–[61].

4.4.3 Hunting

Grey wolves have an excellent sense of smell which enables them to locate the position of prey, chase and encircle the prey. The hunt is usually guided by the alpha. Occasionally, the beta and delta also take part in the hunt. It is, however, apparent that from the inherent capability of the wolves, the alpha, beta and delta have better knowledge about the prospective location of prey than the Omega. Consequently, the first three solutions are considered optimal and the other search agents should update their positions according to the position of the optimal search agents. For this purpose, the following equations are formulated to assume the hunting strategy of the wolves and find the propitious regions for each of the best search agents [59]–[61].

$$D_\alpha = |C_1 X_\alpha(t) - X(t)| \tag{50}$$

$$D_\beta = \left|C_2 X_\beta(t) - X(t)\right| \tag{51}$$

$$D_\delta = |C_3 X_\delta(t) - X(t)| \quad \textbf{(52)}$$

$$X_1 = X_\alpha(t) - A_1 * (D_\alpha) \quad \textbf{(53)}$$

$$X_2 = X_\beta(t) - A_2 * (D_\beta) \quad \textbf{(54)}$$

$$X_3 = X_\delta(t) - A_3 * (D_\delta) \quad \textbf{(55)}$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad \textbf{(56)}$$

Where t denotes the current iteration, $X_\alpha(t), X_\beta(t)$ and $X_\delta(t)$ indicate the position of the grey wolves α, β and δ at the t^{th} iteration, $X(t)$ denotes the position of the grey wolf at t^{th} iteration.

Table 1
Initial parameters of the GWO

Parameter	Value
$\vec{a}$	Linearly decrease from 2 - 0
Population size	100
Maximum generation	250

4.4.4 Attacking prey (Exploitation)

After the wolves have encircled the prey for capturing, the final stage of the hunt is to attack the prey. It worth noting that approaching the prey decreases the value of $\vec{a}$ which in effect decreases the coefficient vector A. This means that A is random variable in the interval [-2$\vec{a}$, 2$\vec{a}$] where $\vec{a}$ is decreased linearly over the iteration cycles. The exploitation of the GWO algorithm starts when $|A| < 1$. However, when the random variable A assumes values in the interval $[-1,1]$, then the next position of a search agent can be random within the position of the prey and the current position of the search agent. This enables the search agents to converge towards an optimal position of prey dictated by the best three solutions (alpha, beta and delta) [59]–[61].

4.4.5 Search for prey (exploration)

Instinctively, grey wolves search for prey in the predatory (solution) space according to the position of the alpha, beta and delta which always move towards the direction of the optimal solution. They tend to move independently to search for prey and converge to attack prey. More importantly, the search (exploration) is determined by the magnitude of the coefficient vector A with random values greater than 1 i.e $|A| > 1$ and that forces the search agent to move away from the prey. Moreover, the coefficient Vector C which takes random values in $[0, 2]$ provides random weights for the prey simply in order to emphasize for $C > 1$ or deemphasize for $C < 1$ the influence of the prey in defining the metric in equation (46). This

randomization ensures exploration and local minima avoidance throughout the optimization process [59]–[61].

4.6 modelling using Time Delay Neural Network TDNN

The prediction models were implemented using Time Delay Neural Networks (TDNN). Unlike the conventional multi-layered neural network architecture, TDNN has a delay line tapped at the input. In essence, such a network has its input neurons modified by introducing a delay line. The inputs of such neurons will now be multiplied by several weights, one for each delay and one for the undelayed input. This modification, however, allows the model to capture the dynamics of the modelled process. [62][63]. In this work a three layer TDNN has been modelled. The three layer TDNN shown in Figure 8 has one hidden layer of ten neurons. Each neuron in the hidden layer has hyperbolic tangent function as the nonlinear function and the output neuron has a pure linear function. The hyperbolic tangent function was chosen as the activation of the hidden layer simply because of its excellent mathematical properties. The architecture of the model is thus 2*L* 10*N* 1L, where *L* represents linear processing neuron and N indicates nonlinear processing neurons. This TDNN architecture is used to model the non-linear temporal dependencies in the acoustic feature vector pattern for the estimation and prediction of the sand flow rate, pressure drop, gas velocity and sand concentration.

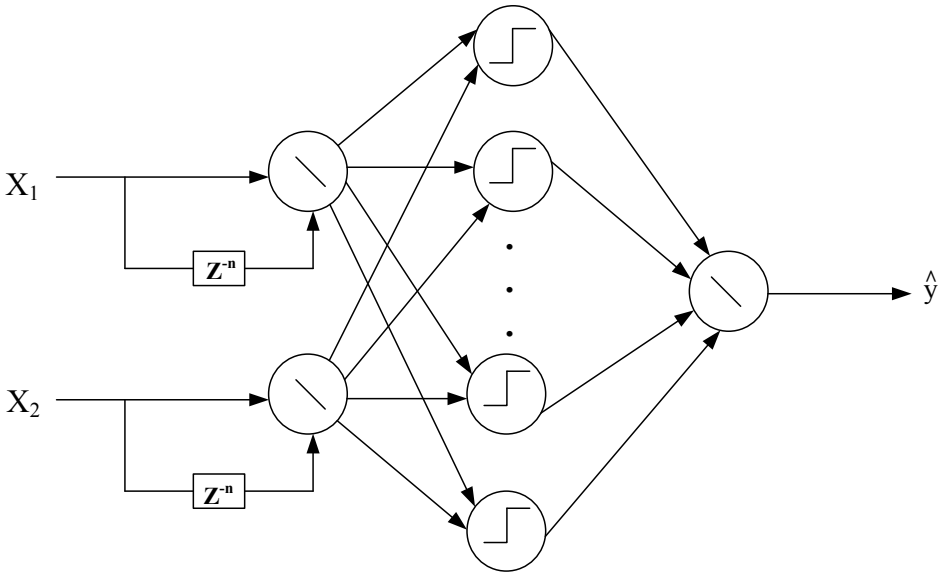


Figure 8: A 3 layer TDNN model

The back propagation algorithm is used to update the weights of the network and the LMA iteratively adapts its damping to minimize the sum of squares cost function. However, there is a caveat relative to backpropagation implementation on the TDNN units. In here, instead of changing the weights on time-shifted connections separately, the weight update is achieved by the average of all corresponding time-delayed weight changes. In this way, the network is compelled to ascertain meaningful input pattern regardless of misalignment of the pattern in time. This is, therefore, a significant property, as it makes the network independent of error-prone preprocessing algorithms that otherwise would be needed for time alignment [63].

Table 2
Database used for the signal processing experiment

SC No	$\dot{m}_g$ (kg/s)	dp (bar)	$\dot{m}_s$ (kg/s)	SC	V_g (m/s)
1	0.0552	1.0847	2.4	43.7	9.9
2	0.0623	1.0620	2.2	35.4	11.3
3	0.0640	0.9887	1.8	28.9	12.1
4	0.0551	0.4412	0.6	11.6	13.9
5	0.0653	0.2077	0.2	3.0	18.0

5 Results and discussion

In the following subsections, we will demonstrate the results of the analysis for the acoustic signals generated when the solid particles hit the bend component of the pipeline using the various signal processing techniques described in the previous sections. For simplicity in the analysis, five instances of target events belonging to 5 different concentration acoustics were considered from the database as shown in Table 2. Additionally, the results of the performance evaluation for the recognition of the solid characteristics data in the pipeline which involves two different methods are also presented. These methods are compared in terms of identification accuracy described by NRMSE.

5.1 Acoustic signal processing results

This section presents the analysis of the raw acoustic signal generated by the impacting solid particles. A Vstatistics analysis results of the acoustic signals, shown in Figure 9, reveals the nature and variability property of the signals overtime, the broad peak at around lg(n) = 3. This broad peak indicates the tendency of the signals to reverse their trends over the time window considered. Figure 9 presents the effects of solids concentration on the energy distributions of the IMFs of acoustic signals. It can be seen in Figure 9 that the acoustic fluctuation energy is mainly distributed at the scale of IMF1 for all the solids concentrations. A possible explanation for these results may be the impacting particles are in suspension flow. However, the energy distribution shifts towards the lower frequency IMFs particularly for solid concentrations: 43.7; 28.9 and 11.6 respectively. As a result, Figure 10 exhibit increased energy distribution fraction in the scales of IMF10 ~ IMF16. Probably, this is due to the slight tendency for change in the flow mechanism of the impinging solid particles in the pipeline. Whether this energy distribution shift is actually related to the change in flow mechanism deserves further investigation.

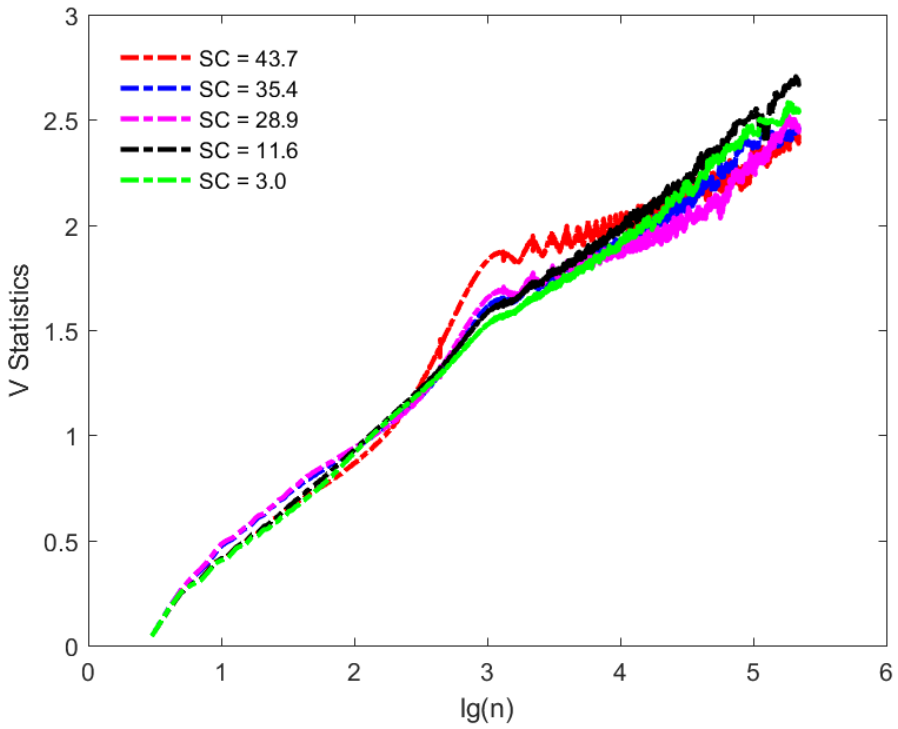


Figure 9 Vstatistics plots of the acoustic signals

Table 3
Concentration of the signals under analysis

Concentration number	**Solid Concentration (SC)**
1	43.7
2	35.4
3	28.9
4	11.6
5	3.0

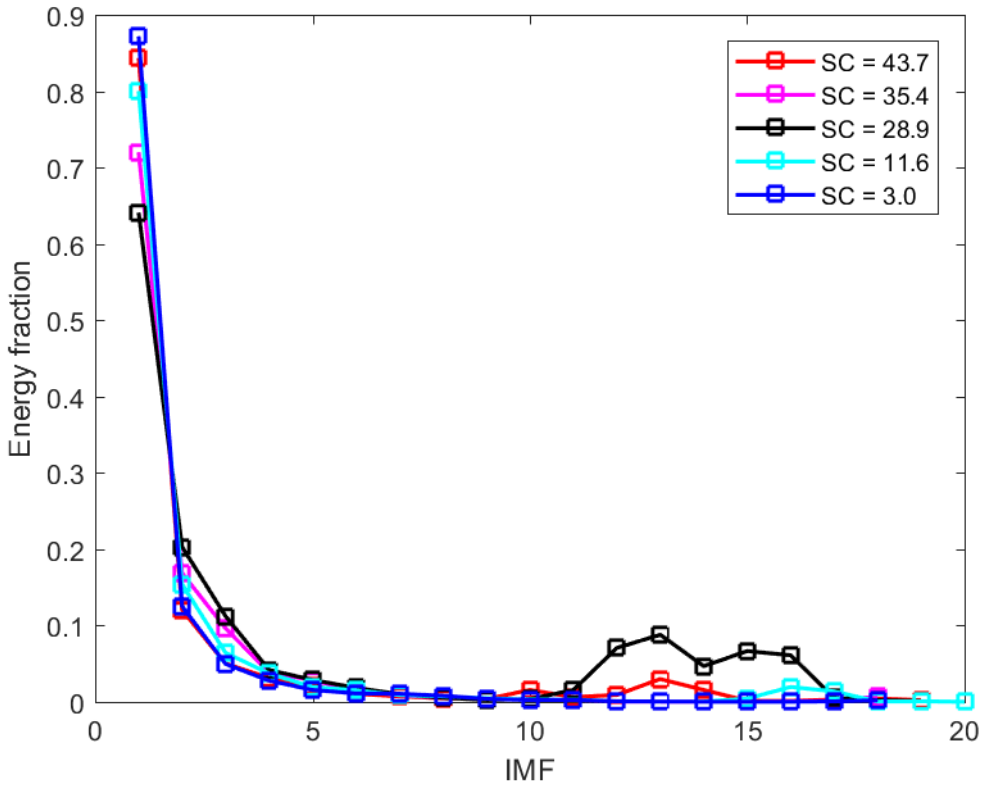


Figure 10 effects of solids concentration on the energy distributions of the IMFs of acoustic signals

Figure 11 presents the spectral power of the acoustic signals for each of the five solid concentrations shown in Table 3. From this figure, one can immediately see that all the acoustic signals show significant power concentration around 1 – 1.3 KHz. This peakiness of the spectra at a relatively wide frequency range can be as a result of the solid flow. However, the details in the low-frequency region are unclear due to the approximately higher magnitude. Additionally, a prevailing band of energies can be easily seen in Figure 10. All characteristics of the spectra are quite significant. Collectively, the energies in the overall frequency range contribute to the spectral power feature. Furthermore, all of the spectrums have substantially different energy distribution patterns caused by the difference in the solids concentrations in the pipeline.

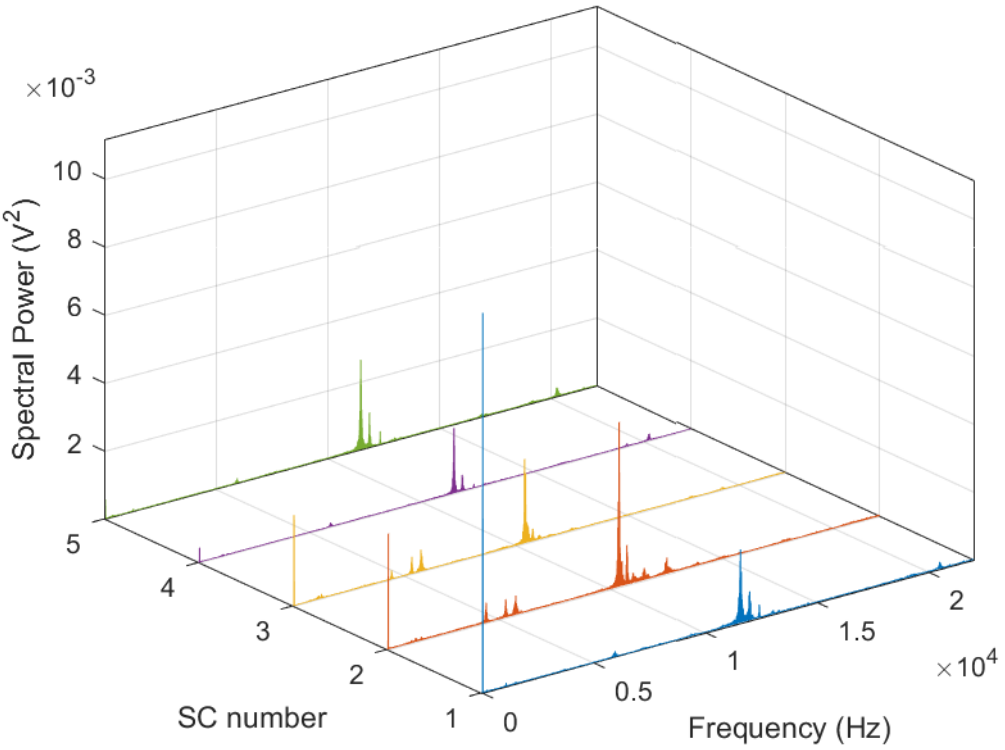


Figure11 Spectral power plots of the acoustic signals for solid concentrations 43.7, 35.4, 28.9, 11.6 and 3.0 over the bandwidth 22.05 KHz.

Figure 12 gives the Histograms of the mean for energy segments of the solid concentrations signals from which the energy fluctuations in the different signals are clearly shown in this probability density – feature value distribution. Besides the clear energy fluctuation of the acoustic signals, there is one more interesting new observation. The fluctuations of the signals energy in relation to the solid concentration are non-linear. The more likely explanation is that the pressurisation in the sand feeder is simply not being regular at the pre-set level, which has significant influence on the signals strength. This figure indicates that the values of this summary statistic are indeed higher for the solid concentration, SC = 35.4. In summary, there is no alternative to the energy measure in terms of robustness to other different kinds of recording variabilities.

The Histograms of the mean of ZCR over the short-term frames of the concentration segments are shown in figure 13. As expected, the signal of the solid concentration, SC = 3.0 contains relatively high zero crossing rate values compared with the trends in the other concentration signals, as shown in Figure 13. This is because the signal of SC = 3.0 is apparently dominated by low amplitude values when compared with other signals. As discussed previously, zero crossing rate of the signals indicate the measure of their amount of randomness. Additionally, the ZCR value as a guide, the randomness in the SC = 3.0 signal can be seen to have one-to-one correspondence with its ZCR value. This observation is indeed consistent with the higher ZCR values as in figure 12. It could be concluded that the lower the concentration, the greater is the amount of randomness in the signals under analysis.

The results of the histograms on the energy entropy for the concentration signals under consideration are shown in Figure 14. Overall, the histograms clearly show the signals exhibit relatively high energy entropy values though the differences in these values are still clear. Interestingly, based on the idea of

the high energy entropy values, one can understand that the signals contain no sudden change in their energies. This is a clear indication that the acoustic measurements are taken under stable flow conditions. Yet, these measurements represent a physical phenomenon, and as such treated by the energy entropy, again a clear indication that the energy entropy can be used to analyse the signals.

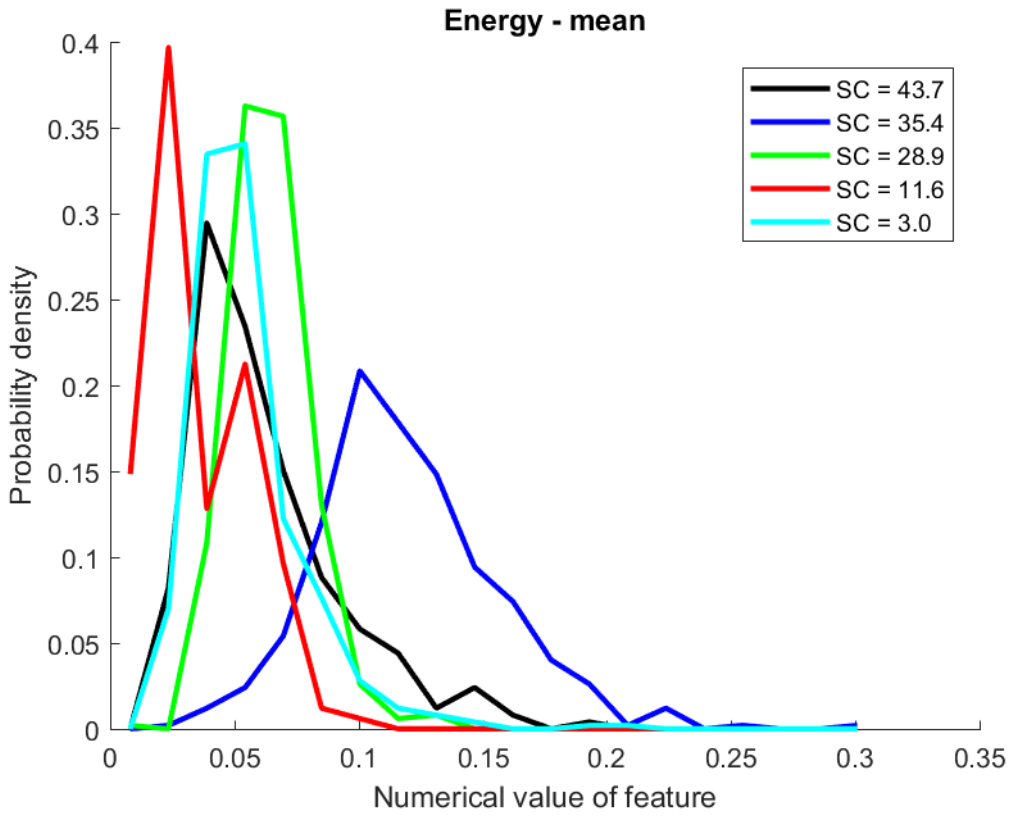


Figure 12 Histograms of the mean of the short-term energy for the five solids concentrations

In Figure 15, the histograms of the mean value of the spectral centroid for acoustic segments from the five solids concentrations are presented. It can be seen that the SC = 3.0 is clearly shown with the expected spectral centroid values. This implies the spectrum of the signal for that concentration is inherently dominated by high frequencies. The only unexpected component of the results is the higher spectral centroid values for the SC = 43.7 signal. It represents the suspension of the solid particles in the medium. On one hand the respective values for this statistic are relatively lower for the other signals. However, taken the uncertainty of gas – solid flow behaviours into consideration, these findings are therefore significant.

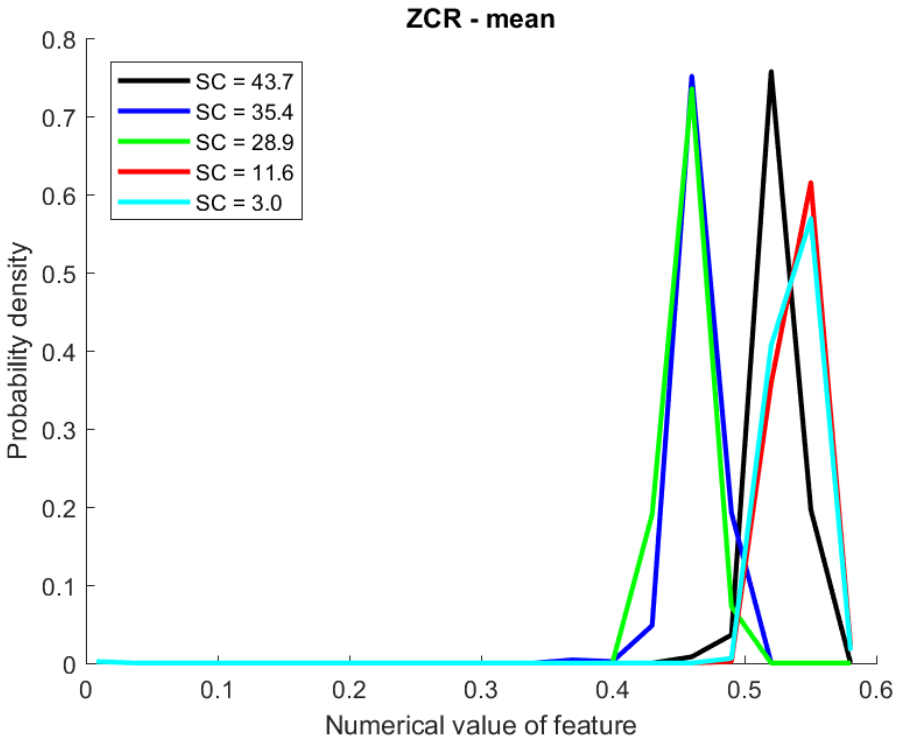


Figure 13 Histograms of the mean value of the sequence of values of the Zero Crossing Rate (ZCR) for the five solids concentrations

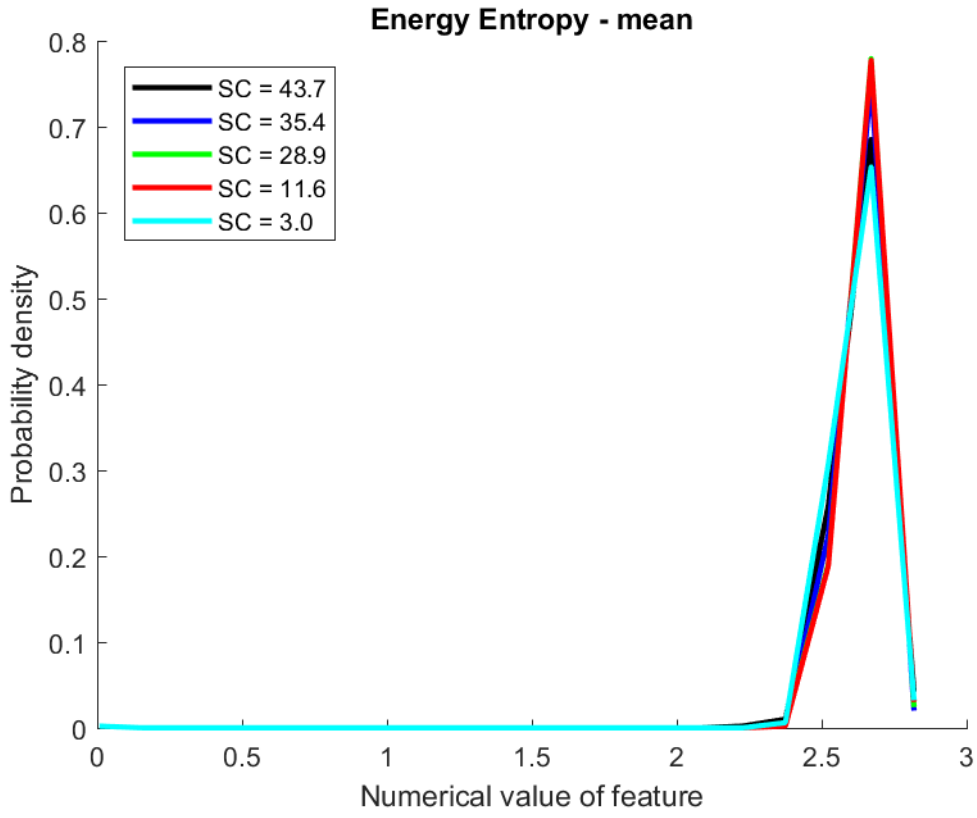


Figure 14 Histograms of the mean value of the entropy of energy for acoustic segments from the five solids concentrations

Applying the spectral spread method to the signals, we have the histograms of the mean value of it for the acoustic segments as shown in Figure 16. Interestingly, this metric measures the dispersion of its values around its spectral centroid. Seemingly low values of the spectral centroid correspond to signals whose spectrum is tightly concentrated around its centre of gravity. For these signals under analysis, the results in Figure 16 indicate that the spectrograms portions of the SC = 43.7 and SC = 35.4 signals are more tightly spread around their centroid than the other signals. It should be noted, however, that this observation is not the same for every SC = 43.7 and SC = 35.4 signals as the correlation between the signals and the concentrations is not linear and constant.

Apparently, the spectral entropy is a metric designed to capture the uniformity in the signal spectrum. Although the results of the histograms of the mean of sequences of spectral entropy of segments from the five signals under analysis, given in Figure 17 show the variation of probability as a function of the spectral entropy values, it is not easy to explain the trends in terms of solids interaction with the bend of the pipeline. In general, the spectral entropy value is low for a spectrum with flat distribution whereas spectrum which contain sharp peaks exhibit higher values. As shown in Figure 17, the SC = 3.0 signal yields the lowest spectral entropy value among the concentrations. The reason seems clear: there is no so much variation in the spectrum in the signal. The values of this statistic for the spectral entropy of SC = 35.4 and SC = 28.9 stay almost identical almost to that of SC = 11.6 at high spectral value. Likewise, the trend for SC = 43.7 and SC = 3.0 but at relatively lower spectral value. Overall, these high values in the spectral entropy indicate the randomness in the spectrum of the signals.

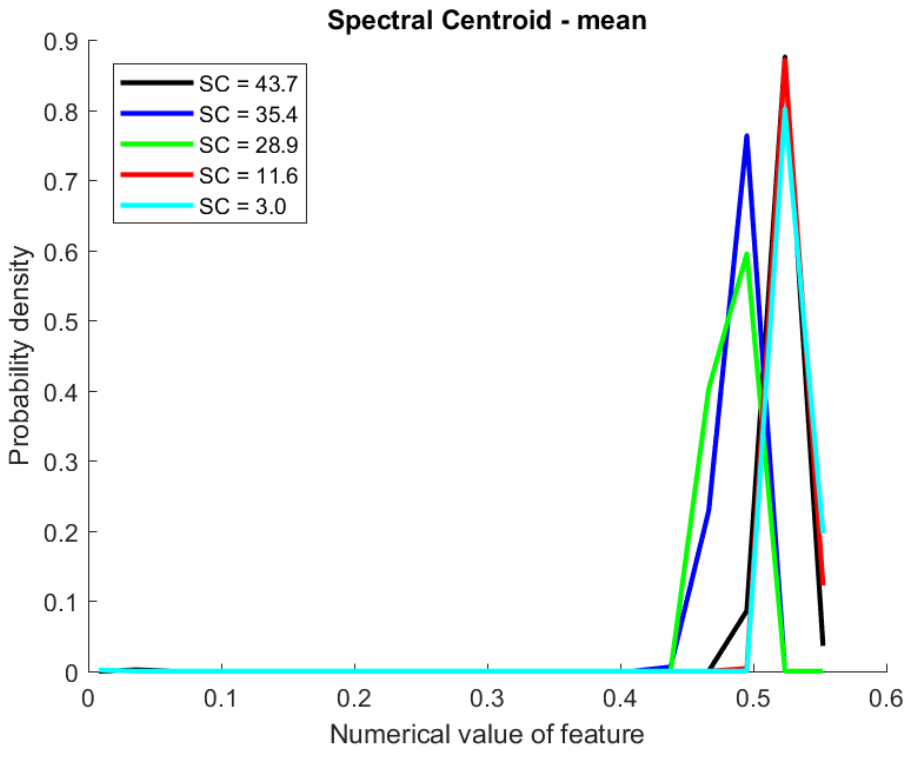


Figure 15 Histograms of the mean value of the sequence of values of the spectral centroid for the five solid concentration signals under consideration

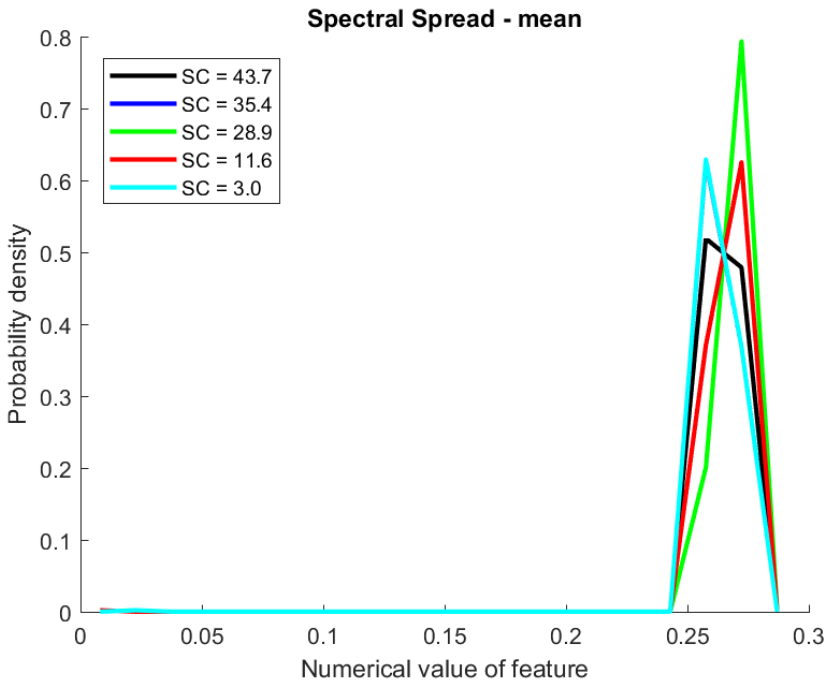


Figure 16 Histograms of the mean value of the sequences of the spectral spread feature for the five solid concentration signals

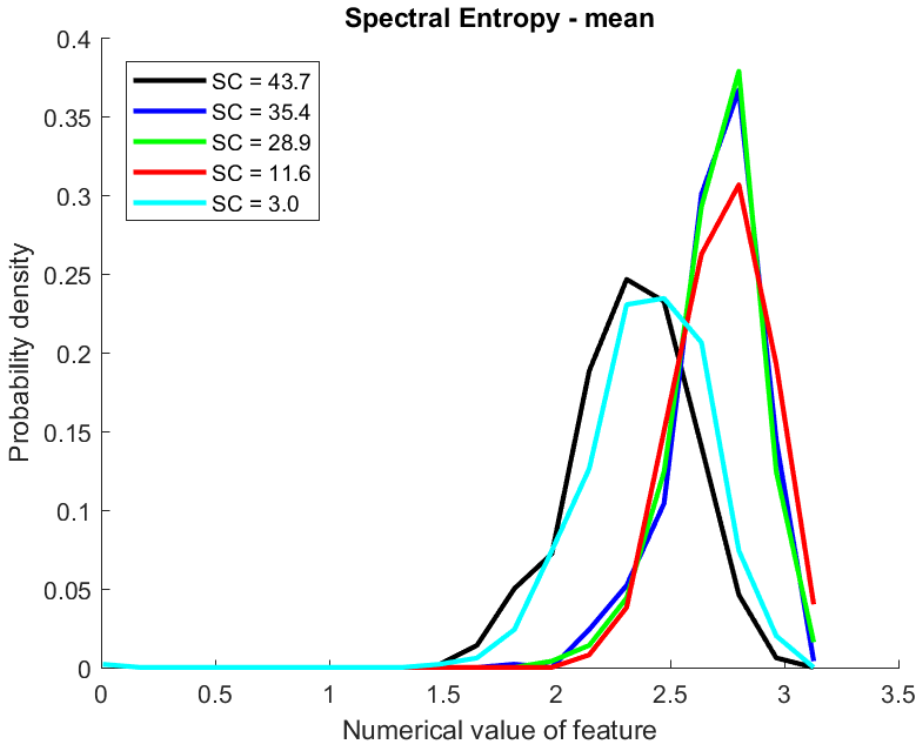
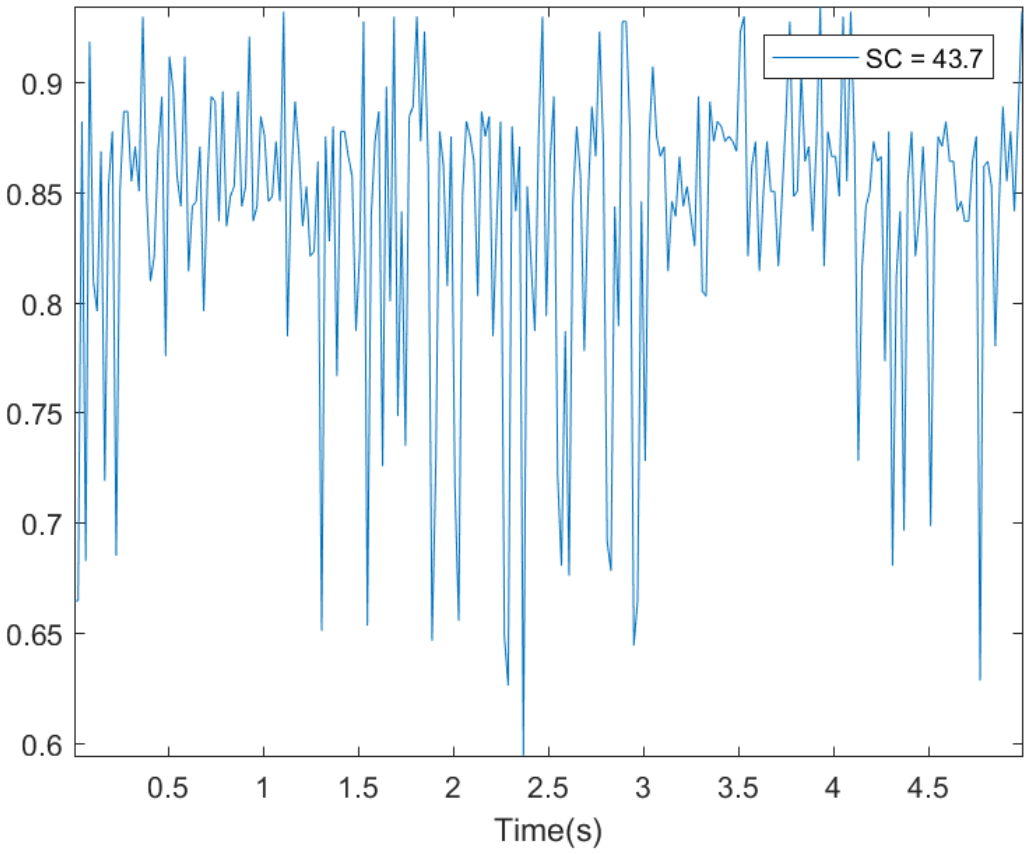
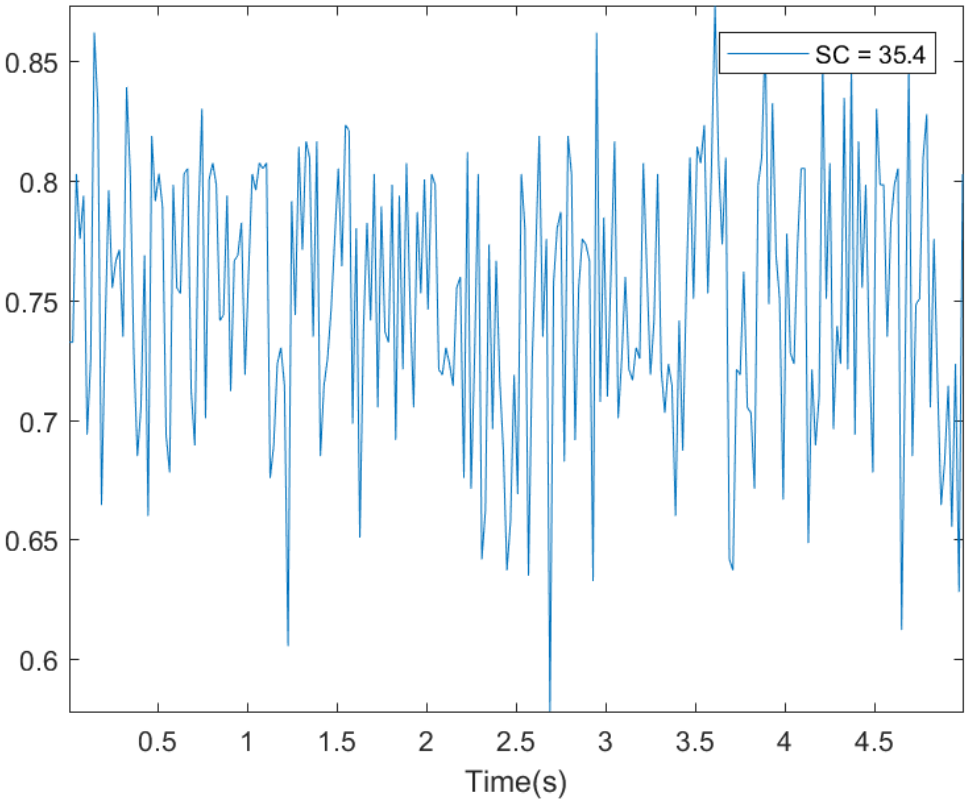
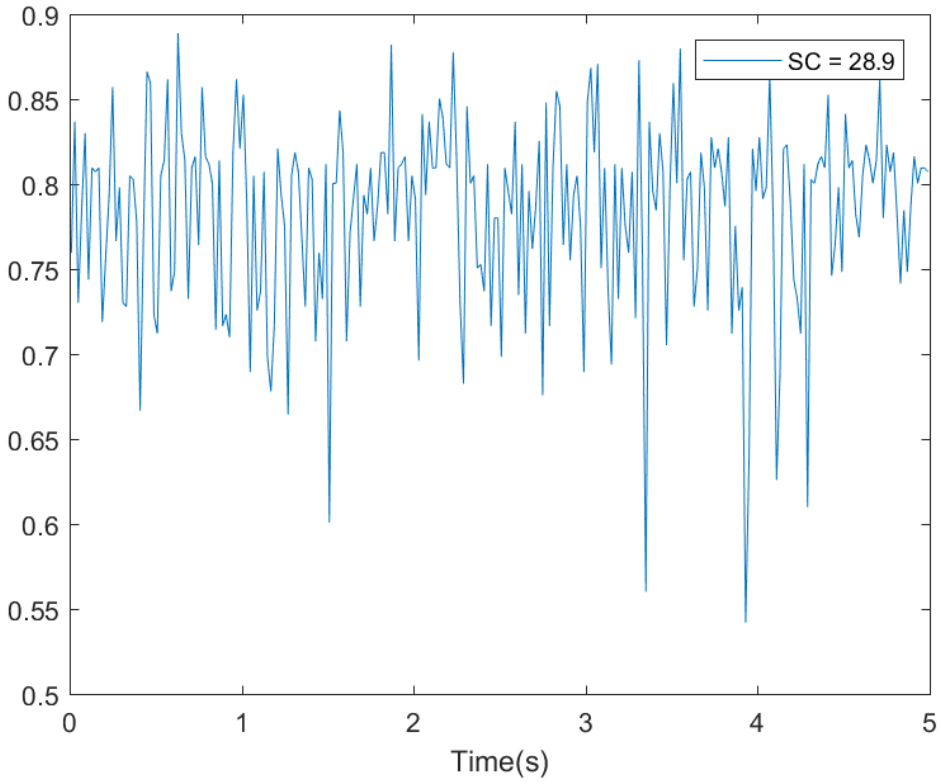


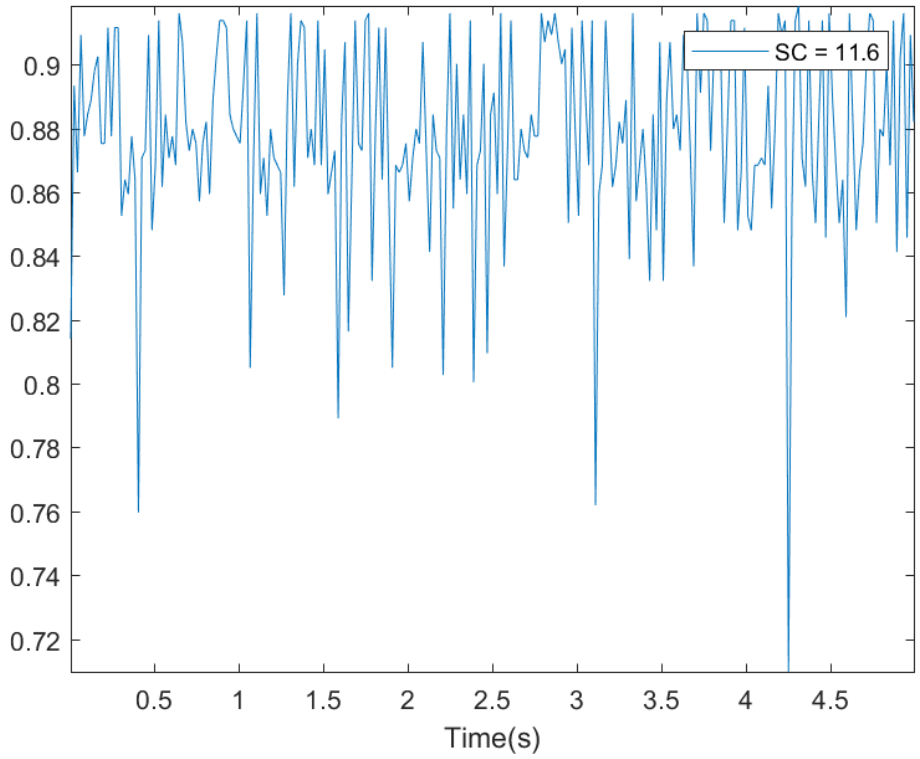
Figure 17 Histograms of the mean of sequences of spectral entropy of segments from the five signals under analysis

Figure 18 presents the sequence of spectral flux for the five concentration acoustic signals under analysis. It can be seen that all the signals exhibit low spectral flux values regardless of the amount of solid concentration in the pipeline. These suggest that all the signals have slowly varying or nearly constant spectral properties. This conception of low spectral flux values clearly makes physical sense from the nature of the interactions of the solid particles with the bend component of the pipeline which are random. Thus, in this case, the details about the variability in spectral properties of the signals have been determined quantitatively.




SC = 35.4
0.85
0.8
0.75
0.7
0.65
0.6
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Time(s)


SC = 28.9
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0
1
2
3
4
5
Time(s)



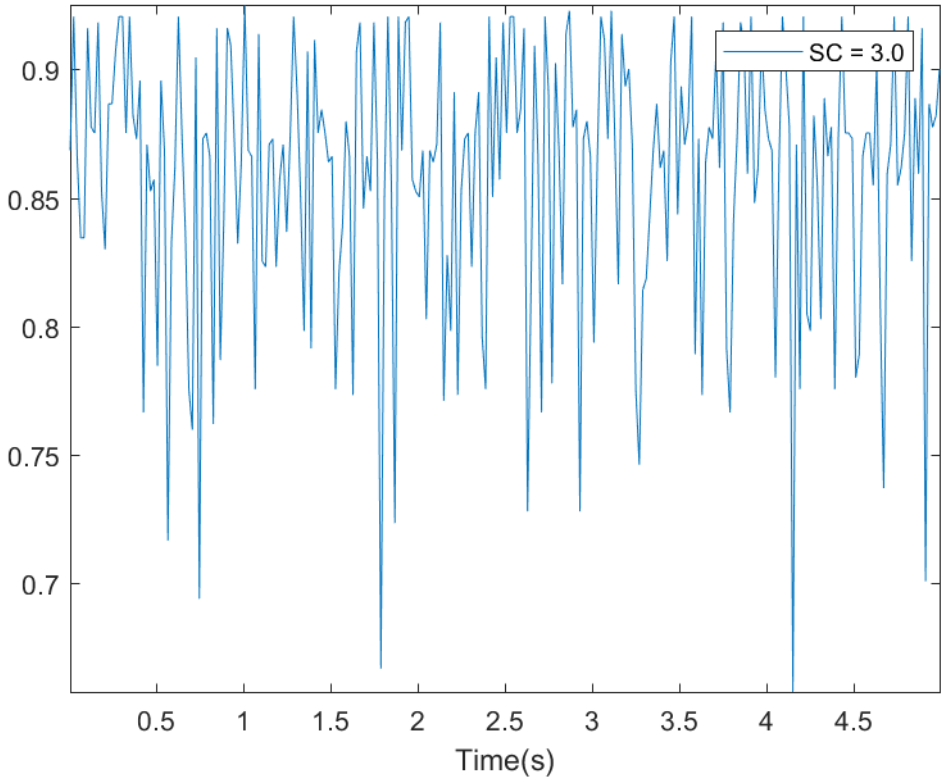


Figure 18 Sequence of spectral flux for the five concentration acoustic signals under analysis: a) SC = 43.7b) SC = 35.4 c) SC = 28.9 d) SC = 11.6 e) SC = 3.0

Figure 19 gives the histograms of the mean value of sequences of spectral rolloff values. From this figure, it can be seen that the characteristics of the signals (SC = 43.7, SC = 11.6 and SC = 3.0) tend to higher values for this feature. This is an indication that the signals contain higher frequencies compared to the other signals. As such 90% of the signals energy lies in the high frequency region. In this realm, the interpretation of this feature serves to give more light about the flow conditions in the pipeline. It can resolve the suspension flow mechanism in the pipeline provided the value is high.

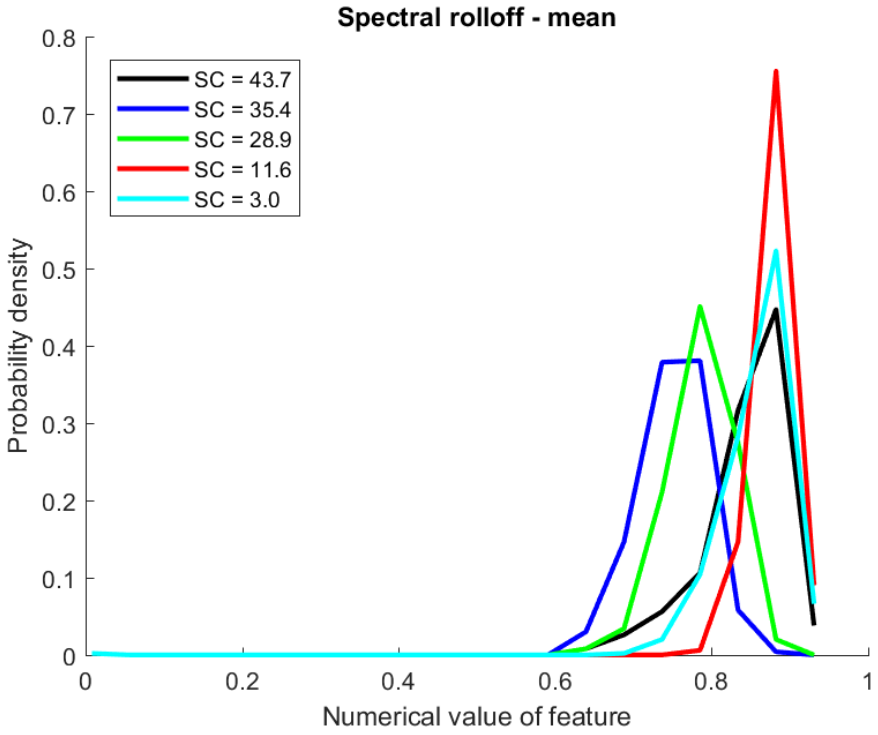


Figure 19 Histograms of the mean of the segments of the spectral rolloff feature for the five solids concentrations signals

Figure 20 shows the histograms of the mean of the 2nd MFCC for the five concentration signals under consideration. The second MFCC index approximates the broadness of a signal spectrum which is in turn related to the spectral centroid. From Figure 19, it can be easily seen that the SC = 35.4 signal gives the highest value for this feature. To explore the differences among the various MFCC indexes, a 2D plot of the first 13 MFCC indexes which describe the coarse spectral envelope of the signals is given in Figure 21. From Figure 21, we can observe the minute variations in the five plots. For example, the SC = 43.7 signal has much higher value in the fourth MFCC than the other signals. However, under consideration of these features all the five concentrations are different from each other.

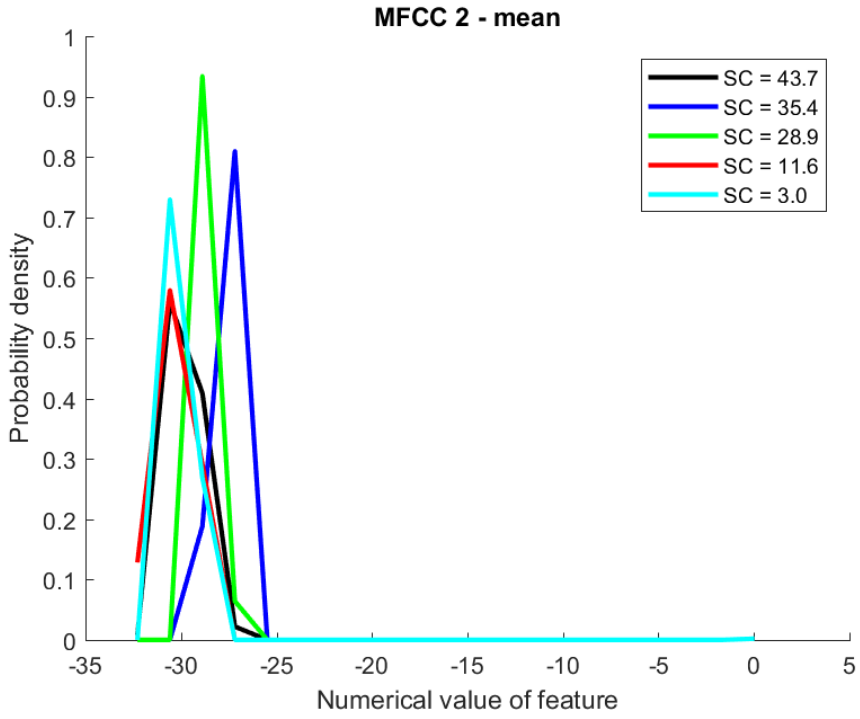
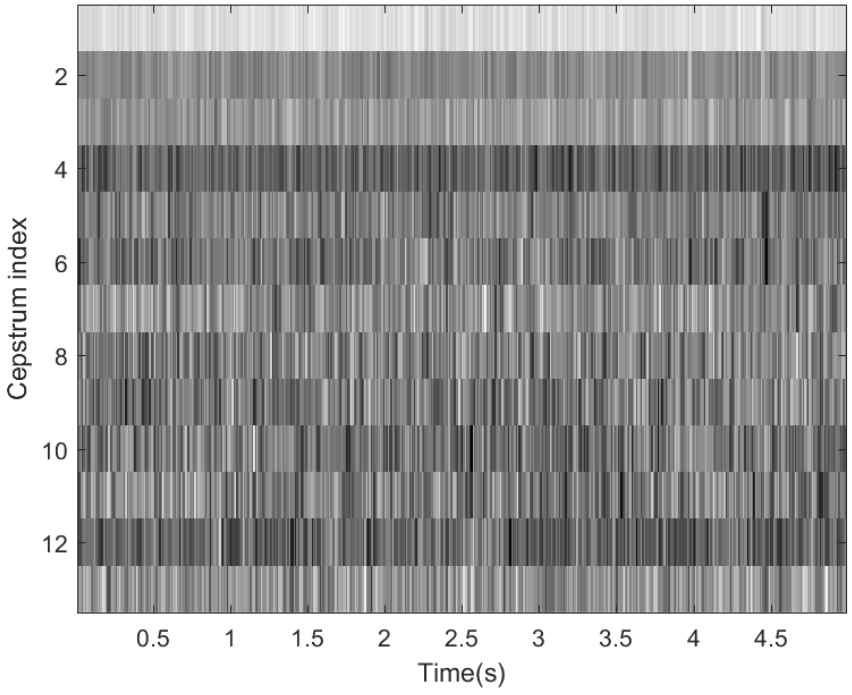
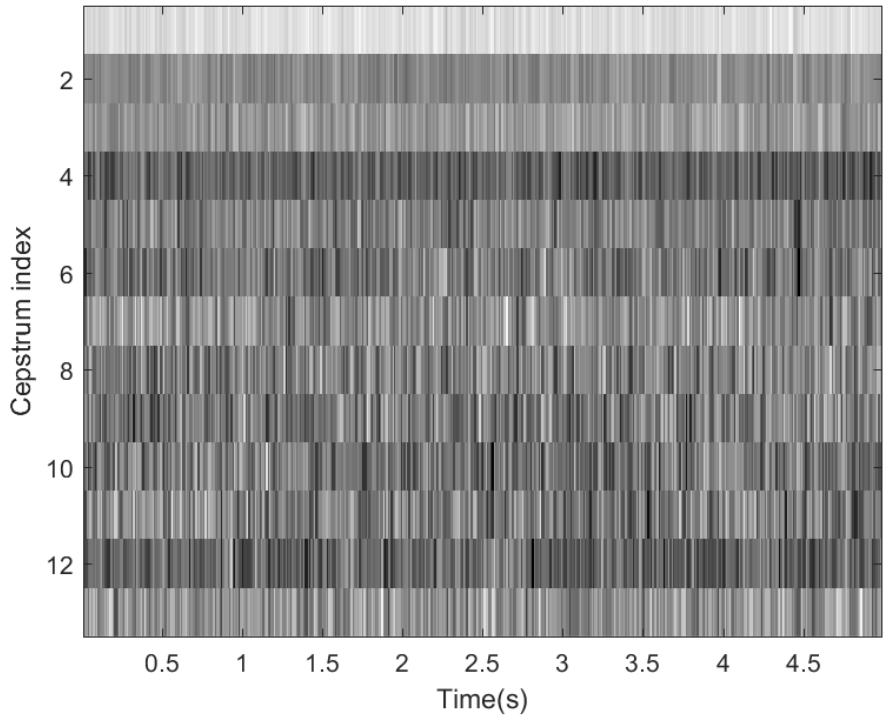
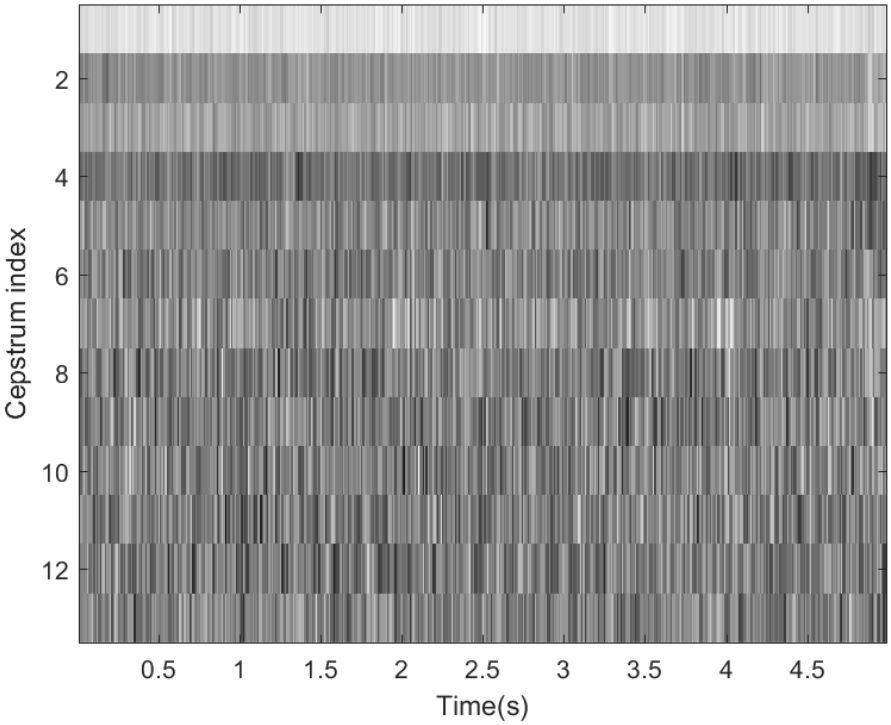


Figure 20Histograms of the mean of the 2nd MFCC for the five concentration signals under analysis




Cepstrum index
2
4
6
8
10
12
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Time(s)


Cepstrum index
2
4
6
8
10
12
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Time(s)

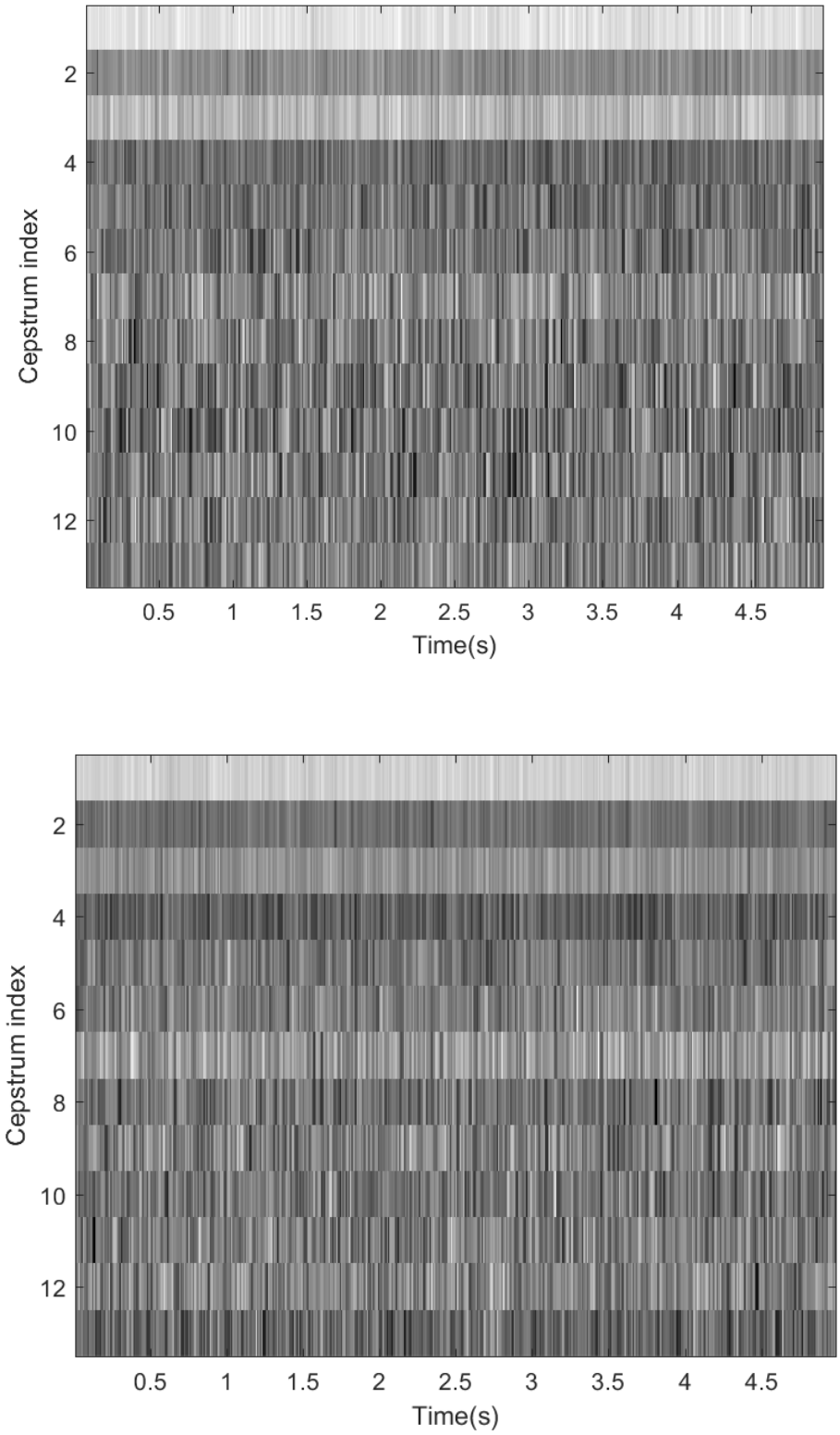


Figure 21 MFCCs for the five solid concentration signals: a) SC = 43.7 b) SC = 35.4 c) SC = 28.9 d) SC = 11.6 e) SC = 3.0

Interestingly, with these idealized features, we have developed the signal processing techniques for the analysis of the concentration signals. Next, we will determine the subsets of these features that are more significant in developing the quantitative recognition models.

5.1.1 Feature Selection

Central to the feature selection processes is to construct a subset of features from the feature vector extracted that are by far more relevant to the recognition system. In fact, the selected features should capture the important characteristics of different kinds of the acoustic signals. The feature selection processes can be categorized as wrapper methods (model-dependent) and filter techniques (model-independent)[64]. In this paper, the wrapper methods have been chosen, particularly in the form of neighbourhood component analysis (NCA). This method indicates how significant features are alone, which gives guidance on how to compose the subset feature vector. The subset of each feature is evaluated by a 5-fold cross validation on the training set. Weights are attached for each feature and these indicate what features to use, where a weight above zero implies that the feature is significant.

Results from the NCA analysis had identify 9 features as the most relevant features from the feature vector, in which 2 features are selected to build the prediction models. The selected features are spectral centroid and spectral entropy.

5.2 Experiment and evaluation performance

In order to evaluate the used prediction models, experiments have been conducted. Firstly the selected features computed from the training dataset were used to train the conventional neural network model described in section 4.4 and the TDNN model described in section 4.6. Accordingly, the test dataset was used to measure the performance. The performance was measured in terms of the normalized root mean square error (RMSE) in the following:

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{57}$$

where y_{max} denotes the maximum value of the actual response, y_{min} represents the minimum value of the actual response and RMSE is the error indicator, defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2} \tag{58}$$

where n is the total number of test trials, y is the actual response and $\hat{y}$ is the predicted or model response. The computer facilities for the experiments are a 3.0 GHz intel (R) core (TM) i5 – 4590S processor, windows 7, 64 – bit operating system, a 8.0 GHz RAM, MATLAB 2017a. Initial parameters for the Grey Wolf Optimiser (GWO) are listed in Table 1, in which the vector $\vec{a}$ varies linearly from 2 to 0 and the population size is 100. The maximum generation number is 250, as it serves as the termination condition for the training.

5.2.1 Experimental results

In order to test the efficiency of the two models with the selected features, these models are evaluated individually. The results presented are in terms of the normalized root mean square error (NRMSE) and are tabulated in Table 2. From the performance results on Table 2, we see that the highest performance (result exhibited by low NRMSE values) is obtained with the TDNN models. This

observation indicates that the TDNN model is particularly suitable for the flow pattern recognition in a particle laden flowline.

Table 4
Performance of two methods for four different solid characteristics models

Method	Model	NRMSE
Classical ANN	PD	0.26
	SFR	0.66
	GV	0.46
	SC	0.29
TDNN	PD	0.20
	SFR	0.18
	GV	0.16
	SC	0.17

PD: Pressure Drop, SFR: Sand Flow Rate, GV: Gas Velocity, SC: Solid Concentration

6 Discussion

In this study, we have presented the techniques of signal processing for the analysis of acoustic signals generated when solid particles strike the bend component of a gas flowline and serve to provide the potential of online quantitative monitoring of the particles in the flowline. The techniques have proven to be versatile and robust in the analysis of the acoustic signals. Additionally, to the authors' knowledge, this is the first study of its kind to use acoustic signals for monitoring the presence of solid particles in a gas flowline quantitatively. Although the use of these signals for qualitative monitoring have been studied previously, quantitative readings about the solid characteristics may be a preferred alternative which can provide a more reliable data that would significantly impact on the economy of the project.

Acoustic solid monitoring has been researched particularly in the oil and gas industry based on the impingements of the solid particles with the wall of the bend through a change of the signal pressure level that is related to the energy of the impact. Effectively, when compared with the velocity, often result to a trend of increasing or decreasing solid particles in the flowline. In addition to the change in signal energy, however, in this study, the change in acoustic that can be heard by the human ear is taken into account. Although the acoustic does not have difference which is perceptible to human hearing, a significant difference was much more pronounced in majority of the spectral properties extracted between the concentration signals.it is, however, evident from the research findings that the signal behaviour differ from concentration to concentration. Additionally, a frequency band of 1 – 1.3 KHz has been identified from the spectra as being the richest with solid concentration related information in this particular monitoring application.

From the V-statistics analysis, it was observed that the signals exhibit non-periodic cycles. The consequence is the misleading energy- frequency distribution when the signals are analysed with the conventional methods of processing. However, the results from our developed methods show good grounds on the front of solid concentration identification in the pipeline. Recognition of the flow conditions in the pipeline has the potential to, among other things, allow the flow to be maximised and therefore preventing risk of unscheduled production deferrals. Hence, this approach provides a very important step towards bridging the gap between existing techniques, which are capable of providing

qualitative readings, and the requirement of quantitative monitoring. This is especially important in the critical area of flow assurance and production optimisation. The potential for solid particles in the flowlines to accelerate erosion damage to major infrastructures can result in unacceptable safety and environment issues as well as economic concerns. Hence, the focus of future research will be on the generated acoustic signal behaviour with mechanical wear formation at the bend, since the prediction of wear formation will be the primary intention of any future condition monitoring system. The analysis of the sensor outputs with the developed methods reveals the complexity of the gas – solid flow. This is because there is no clear pattern between the features extracted from the methods and the flow conditions.

A classical multi-layered artificial neural network quantitative model was trained on a subset of the features extracted on the acoustic emitted by the solid concentration in the flowline. However, this model failed to reveal a significant pattern between the selected features and the solid characteristics data in the flowline. One explanation for this failure may be the inability of the classical neural network architecture to deal properly with dynamic nature of acoustic event. However, a Time-Delay neural network (TDNN) that addresses this aspect of the acoustic event was trained on the acoustic data. Performance evaluation demonstrates superior pattern recognition results has been achieved using this approach.

7 Conclusion

The combination of the signal processing techniques and the machine learning approach has offered a powerful method for online quantitative monitoring of solid particulate materials in a gas flowline. The key to this approach is the signal processing methods, which enables the acoustic data to be reduced into such a form that the machine learning method can easily work with. These methods forming the basis of the feature extraction are computationally efficient. To the author's best knowledge, this is the first time acoustic signal analysis has been quantitatively correlated to the concentration of solids in a gas flowline. This is particularly significant because changes in the acoustic generated from the impacting particles are not perceptible to human hearing. Furthermore, the features resulting from the analysis are further optimised using the neighbourhood component analysis (NCA) for regression and their relevance for the responses have been studied separately. We have found that the spectral centroid and the spectral entropy are the most relevant features to be considered in building the machine learning model. A great advantage of NCA is the improved computational efficiency that could be achieved in developing the model. Rather than the complete feature set, a subset has been considered without affecting the overall performance.

Other than the practical methodology, the most important conceptual innovation of the present research is the consideration for Time-Delay Neural Networks (TDNN) - a variation of classical neural networks prepared to accommodate dynamics in flow mechanism, as the predictive model. The TDNN model has the ability to learn complex nonlinear decision surface. The performance of the TDNN over testing data from the solid characteristics in the flowline has been extensively evaluated. The TDNN achieved low values of NRMSE for all the models compared to the classical neural network models for the same testing data. The power of the TDNN lies in their ability to develop shift invariance for making optimal decisions. In general, this property holds significant promise for flow condition pattern recognition system. This could go a long way in helping to overcome the representational weakness of other conventional recognition systems faced with the uncertainty and variability in real-world data.

Acknowledgement

The authors would like to acknowledge the EPSRC funded ARCHIE-WEST High Performance Computer the results for signal processing were obtained. EPRC grant no. EP/K000586/1.

References

[1] P. Thiruvenkatanathan, T. Langnes, P. Beaumont, D. White, and M. Webster, "Downhole Sand Ingress Detection Using Fibre-Optic Distributed Acoustic Sensors," in *International Petroleum Exhibition & Conference*, 2016, p. 9.

[2] I. McKay, P. Russ, and J. Mohr, "A Sand Management System for Mature Offshore Production Facilities," in *International Petroleum Technology ...*, 2008, no. 12784, pp. 1–8.

[3] B. Dudley, "BP Energy Outlook 2017 edition BP Energy Outlook 2017 edition," 2017.

[4] T. Haugsdal, "The Most Efficient Use of Acoustic Sand Monitors . Lessons Learned from Many Years of Operation Sand Monitoring Acoustic Sand Monitoring Introduction," *Soc. Pet. Eng.*, no. SPE-185888-MS, pp. 1–12, 2017.

[5] A. Gupta *et al.*, "Getting the Best Out of Online Acoustic Sand Monitoring System : A Practical Method for Quantitative Interpretation," in *International Petroleum Technology Conference*, 2016, pp. 1–11.

[6] A. Mackinnon, J. Brown, and G. K. Brown, "Keeping Acoustic Sand Monitoring Simple," *Corros. Conf. Expo*, no. 11396, pp. 1–14, 2011.

[7] M. Odigie, S. A. Shirazi, B. S. McLaury, and S. Cremaschi, "Acoustic Monitor Threshold Limits for Sand Detection in Multiphase Flow Production System," *SPE Int. Conf. Exhib. Oilf. Corros.*, vol. 154378, pp. 1–13, 2012.

[8] M. E. El-Alej, "Monitoring Sand Particle Concentration in Multiphase Flow Using Acoustic Emission Technology," *World Acad. Sci. Eng. Technol.*, vol. 7, no. 6, pp. 1–7, 2014.

[9] D. C. Finfer, V. Mahue, S. V Shatalin, T. R. Parker, and M. Farhadiroushan, "Borehole Flow Monitoring using a Non-intrusive Passive Distributed Acoustic Sensing (DAS)," in *SPE Annual Technical Conference and Exhibition*, 2014, pp. 1–9.

[10] R. C. Guido, "A tutorial on signal energy and its applications," *Neurocomputing*, vol. 179, pp. 264–282, 2016.

[11] S. Krishnan, K. Umapathy, and B. Ghoraani, "Audio signal processing using time-frequency approaches: Coding, classification, fingerprinting, and watermarking," *EURASIP J. Adv. Signal Process.*, vol. 2010, p. 28, 2010.

[12] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, 1998.

[13] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - Part 1: Fundamentals," *Proc. IEEE*, vol. 80, no. 4, pp. 520–538, 1992.

[14] D. Gerhard, "Audio Signal Classification : History and Current Techniques," 2003.

[15] T. Giannakopoulos, *Introduction to AUDIO ANALYSIS : A MATLAB Approach*. 2014.

[16] T. M. Ventura *et al.*, "Audio parameterization with robust frame selection for improved bird identification," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8463–8471, 2015.

[17] D. Mitrović, M. Zeppelzauer, and C. Breiteneder, "Features for Content-Based Audio Retrieval," in *Advances in Computers*, vol. 78, no. 10, 2010, pp. 71–150.

[18] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and*

Applications, Fourth. New Jersey: Pearson Education, 2007.

[19] T. K. Rawat, "Discrete Fourier Transform," in *Digital Signal Processing*, Oxford University Press.

[20] L. Luo, Y. Yan, P. Xie, J. Sun, Y. Xu, and J. Yuan, "Hilbert-Huang transform, Hurst and chaotic analysis based flow regime identification methods for an airlift reactor," *Chem. Eng. J.*, vol. 181–182, pp. 570–580, 2012.

[21] P. Lu, D. Han, R. Jiang, X. Chen, C. Zhao, and G. Zhang, "Experimental study on flow patterns of high-pressure gas-solid flow and Hilbert-Huang transform based analysis," *Exp. Therm. Fluid Sci.*, vol. 51, pp. 174–182, 2013.

[22] P. Ruvolo, I. Fasel, and J. R. Movellan, "A learning approach to hierarchical feature selection and aggregation for audio classification," *Pattern Recognit. Lett.*, vol. 31, no. 12, pp. 1535–1542, 2010.

[23] D. Sun *et al.*, "On-line nonintrusive detection of wood pellets in pneumatic conveying pipelines using vibration and acoustic sensors," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 5, pp. 993–1001, 2014.

[24] W. Shuiping, T. Zhenming, and L. Shiqiang, "Design and Implementation of an Audio Classification System Based on SVM," *Procedia Eng.*, vol. 15, pp. 4031–4035, 2011.

[25] C. S. Ooi, K. P. Seng, L. M. Ang, and L. W. Chew, "A new approach of audio emotion recognition," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 5858–5869, 2014.

[26] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[27] A. Mittal, S. S. Mallick, and P. W. Wypych, "An investigation into pressure fluctuations for fluidized dense-phase pneumatic transport of fine powders," *Powder Technol.*, vol. 277, pp. 163–170, 2015.

[28] P. N. Le, E. Ambikairajah, J. Epps, V. Sethu, and E. H. C. Choi, "Investigation of spectral centroid features for cognitive load classification," *Speech Commun.*, vol. 53, no. 4, pp. 540–551, 2011.

[29] N. Kamarudin, S. A. R. Al-Haddad, S. J. Hashim, M. A. Nematollahi, and A. R. Bin Hassan, "Feature extraction using Spectral Centroid and Mel Frequency Cepstral Coefficient for Quranic Accent Automatic Identification," in *2014 IEEE Student Conference on Research and Development, SCOReD 2014*, 2014, pp. 0–5.

[30] M. García, J. Poza, D. Abásolo, D. Santamarta, and R. Hornero, "Analysis of intracranial pressure signals recorded during infusion studies using the spectral entropy," in *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2013, vol. 2013, pp. 2543–2546.

[31] L. K. Isaacson, "Spectral entropy, empirical entropy and empirical exergy for deterministic boundary-layer structures," *Entropy*, vol. 15, no. 10, pp. 4134–4158, 2013.

[32] A. M. Toh, R. Togneri, and S. Nordholm, "Spectral entropy as speech features for speech recognition," *Comput. Eng.*, no. 1, pp. 22–25, 2005.

[33] J. P. Bello, L. Daudet, S. A. andC. Duxbury, M. Davies, and M. Sandler, "A Tutorial on Onset Detection in Musical Signals," *IEEE Trans. Speech Audio Process. Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, 2005.

[34] J. Xie, M. Towsey, J. Zhang, and P. Roe, “Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms,” *Appl. Acoust.*, vol. 113, pp. 193–201, 2016.

[35] M. Kos, Z. Kačič, and D. Vlaj, “Acoustic classification and segmentation using modified spectral roll-off and variance-based features,” *Digit. Signal Process. A Rev. J.*, vol. 23, no. 2, pp. 659–674, 2013.

[36] T. L. F. Evangelista, T. M. Priolli, C. N. S. Jr, B. a. Angelico, and C. a. a. Kaestner, “Automatic Segmentation of Audio Signals for Bird Species Identification,” *2014 IEEE Int. Symp. Multimed.*, pp. 223–228, 2014.

[37] M. Sahidullah and G. Saha, “Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition,” *Speech Commun.*, vol. 54, no. 4, pp. 543–565, 2012.

[38] J. X. Du, C. M. Zhai, Y. L. Guo, Y. Y. Tang, and P. C. Chun Lung, “Recognizing complex events in real movies by combining audio and video features,” *Neurocomputing*, vol. 137, pp. 85–95, 2014.

[39] S. Sigurdson, K. B. Petersen, and J. Larsen, “Mel Frequency Cepstral Coefficients: An Evaluation of Robustness of MP3 Encoded Music,” in *Proceedings of the 7th International Conference on Music Information Retrieval*, 2006, pp. 3–6.

[40] K. Chakraborty Asmita Talele Savitha Upadhya, “Voice Recognition Using MFCC Algorithm,” *Int. J. Innov. Res. Adv. Eng.*, vol. 1, no. 10, pp. 2349–2163, 2014.

[41] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, 1943.

[42] M. Riedmiller, “Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms,” *Comput. Stand. Interfaces*, vol. 16, no. 3, pp. 265–278, 1994.

[43] V. P. Plagianakos, D. G. Sotiropoulos, and M. N. Vrahatis, “An Improved BP method with adaptative lerarning,” in *Advances in Convex Analysis and Global Optimization*, 2008, pp. 1–5.

[44] V. Kecman, *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models*. Massachussetts Institute of Technology, 2001.

[45] G. D. Magoulas, V. P. and Plagianakos, and M. N. Vrahatis, “Neural network-based colonoscopic diagnosis using on-line learning and differential evolution,” *Appl. Soft Comput.*, vol. 4, no. 4, pp. 369–379, 2004.

[46] V. P. Plagianakos, G. D. Magoulas, and M. N. Vrahatis, “Chapter 2. Learning Rate Adaption in Stochastic Gradient Descent,” in *Advances in Convex Analysis and Global Optimization*, 2008, pp. 15–26.

[47] A. D. Anastasiadis, G. D. Magoulas, and M. N. Vrahatis, “Improved sign- based learning algorithm derived by the composite nonlinear Jacobi process,” *J. Comput. Appl. Math.*, vol. 191, no. 2, pp. 166–178, 2006.

[48] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Muller, “Efficient BackProp,” in *Neural Networks: Tricks of the Trade*, G. B. Orr and K.-R. Muller, Eds. Newyork: Springer-Verlag, 1998, pp. 9–50.

[49] A. K. Jain, J. Mao, and K. M. Mohiuddin, “Artificial Neural Networks: A Tutorial -

Computer," *IEEE journals Mag.*, vol. 29, no. 3, pp. 31–44, 1996.

[50] K. Mannepalli, P. N. Sastry, and M. Suman, "A novel Adaptive Fractional Deep Belief Networks for speaker emotion recognition," *Alexandria Eng. J.*, vol. 56, pp. 485–497, 2017.

[51] V. P. Plagianakos, G. D. Magoulas, and M. N. Vrahatis, "Learning in multilayer perceptrons using global optimization strategies," *Nonlinear Anal. Theory, Methods Appl.*, vol. 47, no. 5, pp. 3431–3436, 2001.

[52] V. P. Plagianakos, G. D. Magoulas, and M. N. Vrahatis, "Supervised Training Using Global Search Methods," in *Advances in Convex Analysis and Global Optimization*, 2013, pp. 3–14.

[53] M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *IEEE Int. Conf. Neural Networks - Conf. Proc.*, vol. 1993–Janua, pp. 586–591, 1993.

[54] A. D. Anastasiadis, G. D. Magoulas, and M. N. Vrahatis, "Sign-based learning schemes for pattern classification," *Pattern Recognit. Lett.*, vol. 26, no. 12, pp. 1926–1936, 2005.

[55] N. K. Treadgold and T. D. Gedeon, "Simulated annealing and weight decay in adaptive learning: The SARPROP algorithm," *IEEE Trans. Neural Networks*, vol. 9, no. 4, pp. 662–668, 1998.

[56] D. B. Fogel, L. J. Fogel, and V. W. Porto, "Evolving neural networks," *Bilogical Cybern.*, vol. 63, no. 6, pp. 487–493, 1990.

[57] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning from Data: A Short Course*. AMLbook.com, 2012.

[58] C. M. Bishop, *Pattern Recognition and Machine Learning*, vol. 4, no. 4. 2006.

[59] S. Mirjalili, S. Mohammad, and A. Lewis, "Advances in Engineering Software Grey Wolf Optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[60] S. Mirjalili, "How effective is the Grey Wolf optimizer in training multi-layer perceptrons," *Appl. Intell.*, vol. 43, no. 1, pp. 150–161, 2015.

[61] Y. Tonce, K. Priyanto, and A. P. System, "Multi Objective Optimal Power Flow To Minimize Losses and Carbon Emission U sing Wolf Algorithm," in *International Seminar on Intelligent Technology and Its Applications*, 2015, pp. 153–158.

[62] J. Wang, I. Tsapakis, and C. Zhong, "A space – time delay neural network model for travel time prediction," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 145–160, 2016.

[63] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and J. K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE TRANSANCTIONS Acoust. SPEECH, SIGNAL Process.*, vol. 37, no. 3, pp. 1–12, 1989.

[64] J. Ludeña-Choez and A. Gallardo-Antolín, "Acoustic Event Classification using spectral band selection and Non-Negative Matrix Factorization-based features," *Expert Syst. Appl.*, vol. 46, pp. 77–86, 2016.